GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2002, 14:29:52 ; Search time 7114.74 Seconds
                                              (without alignments)
                                              5809.054 Million cell updates/sec

Title: US-09-052-089A-8
Perfect score: 1975
Sequence: 1 GGCACGAGGTCCGGTGGAGC..........CAAAAAAAAAAAAAAAAAAA 1975

Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum-Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Database :   GenEmbl:*
                   1: gb_ba:*
                   2: gb_htg:*
                   3: gb_in:*
                   4: gb_om:*
                   5: gb_ov:*
                   6: gb_pat:*
                   7: gb_ph:*
                   8: gb_pl:*
                   9: gb_pr:*
                   10: gb_ro:*
                   11: gb_sts:*
                   12: gb_sy:*
                   13: gb_un:*
                   14: gb_vi:*
                   15: em_ba:*
                   16: em_fun:*
                   17: em_hum:*
                   18: em_in:*
                   19: em_mu:*
                   20: em_om:*
                   21: em_or:*
                   22: em_ov:*
                   23: em_pat:*
                   24: em_ph:*
                   25: em_pl:*
                   26: em_ro:*
                   27: em_sts:*
                   28: em_un:*
                   29: em_vi:*
                   30: em_htg_hum:*
                   31: em_htg_inv:*
                   32: em_htg_other:*
                   33: em_htgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result   %   Query
No.   Score   Match Length DB   ID   Description
----- -----   ----- ------ --   --   -----------

1   1952   98.8   1981   10   MMU77844   U77844 Mus musculu
2   1880   95.2   2024   10   BC017374   BC017374 Mus muscu
3   1075.4   54.5   2065   6   AR072729   AR072729 Sequence
4   1075.4   54.5   2065   6   BD006990   BD006990 Modulator
5   1072.2   54.3   2002   9   BC019283   BC019283 Homo sapi
6   1072.2   54.3   2010   9   BC000310   BC000310 Homo sapi
7   1062.8   53.8   2007   9   HSU77845   U77845 Human hTRIP
8   952.6   48.2   3694   10   BC006929   BC006929 Mus muscu
9   952.6   48.2   3694   10   AX350344   AX350344 Sequence
10   111.4   5.6   7542   9   AC099330   AC099330 Homo sapi
11   111.4   5.6   163424 2   AL031432   AL031432 Human DNA
12   108.2   5.5   1297747 9   HS465N24   Continuation (13 of
13   101.8   5.2   1110000 2   AC099330_2   Continuation (17 o
14   101.8   5.2   163424 2   AC099330   Continuation (13 of
15   72.4   3.7   39410   5   A7010317   Fugu rubr
16   64   3.2   7218   6   I66494   I66494 Sequence 14
17   57.6   2.9   2065   14   AF410847   AF410847 Ovine her
18   55.2   2.8   203984 2   AC067854   AC067854 Homo sapi
19   54.2   2.7   1110000 2   LMFICHR34_16   LMFICHR34_16
20   53.8   2.7   207683 2   AC098712   AC098712 Mus muscu
21   53.2   2.7   216191 2   AC090495   AC090495 Mus muscu
22   51.6   2.6   3127   14   AF305694   AF305694 Kaposi's
23   51.4   2.6   211158 2   AC073609   AC073609 Mus muscu
24   51.2   2.6   1683   9   CHPINVOL   M26514 Chimpanzee
25   51   2.6   133661 14   U93872   U93872 Kaposi's sa
26   51   2.6   219200 10   AL589701   AL589701 Mouse DNA
27   50.4   2.6   3012   14   AF360120   AF360120 Human her
28   50.2   2.5   180385 2   AC007461   AC007461 Homo sapi
29   50   2.5   125020 14   AF429315   AF429315 Kaposi's
30   49.6   2.5   28559   14   AF148805   AF148805 Kaposi's
31   49.2   2.5   140207 2   AC016837   AC016837 Homo sapi
32   49.2   2.5   164310 9   AC019294   AC019294 Homo sapi
33   49.2   2.5   180795 2   AC027104   AC027104 Homo sapi
34   49.2   2.5   227605 2   AC073564   AC073564 Mus muscu
35   49   2.5   282611 2   AL645746   AL645746 Mus muscu
36   48.8   2.5   206497 10   AC084389   AC084389 Mus muscu
37   48.8   2.5   2160   10   AY033497   AY033497 Mus muscu
38   48.8   2.5   2193   10   AB033744   AB033744 Mus muscu
39   48.2   2.4   5894   9   Y00624   Y00624 Acanthamoeb
40   48.2   2.4   241432 10   AL589661   AL589661 Mouse DNA
41   47.2   2.4   202010 2   AC016961   AC016961 Homo sapi
42   47   2.4   35848   2   AC011553   AC011553 Homo sapi
43   47   2.4   172148 3   LMFP696   LMFP696
44   47   2.4   175463 2   AC095873   AC095873 Rattus no
45   46.8   2.4   199508 2   AC098555   AC098555 Rattus no
46.8 2.4   2931   14   AF192756   AF192756 Kaposi's

ALIGNMENTS

RESULT 1
MMU77844
LOCUS   MMU77844   1981 bp   mRNA   linear   ROD 25-APR-1997
DEFINITION   Mus musculus mTRIP (mTRIP) mRNA, complete cds.
ACCESSION   U77844
VERSION   U77844.1 GI:2039305
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1981)
  AUTHORS   Lee,S.Y., Lee,S.Y. and Choi,Y.
  TITLE   TRAF-interacting protein (TRIP): a novel component of the tumor
             necrosis factor receptor (TNFR)- and CD30-TRAF signaling complexes
             that inhibits TRAF2-mediated NF-kappaB activation
  JOURNAL   J. Exp. Med. 185 (7), 1275-1285 (1997)
   MEDLINE   97258620
REFERENCE   2 (bases 1 to 1981)
  AUTHORS   Lee,S.Y., Lee,S.Y. and Choi,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-Nov-1996) Immunology, The Rockefeller University,
             1230 York Avenue Box 295, New York, NY 10021, USA

```
FEATURES             Location/Qualifiers
     source          1..1981
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..1981
                     /gene="mTRIP"
     CDS             118..1530
                     /gene="mTRIP"
                     /codon_start=1
                     /product="mTRIP"
                     /protein_id="AAB52994.1"
                     /db_xref="GI:2039306"
                     /translation="MPILSLCTICSDFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSR
                     TCPQCRIQVGKKTIINKLFPDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQA
                     IIDFLRDTLEERNATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLK
                     CKMKTMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAYVCVSLKKEYENLKEARKAT
                     GELADRLKKDLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKLMILQG
                     TLSLPPRTNETVSRLVFESPAPVEMNPRLHQPPFGDEIDLNTTFDVNTPPTQTISGSQ
                     HCLPKKLCLERARSPMQNVLKKVHKVSKPEQSLSLGGQRCVGELDEELAGAFPLFIRN
                     AVLGQKQPNRTTAESSRSSTDVVRIGFDGLGGRTKFIQPRDTTIIRPVPVKSKAKSKQK
                     VRIKTVSSASQPKLDTFLCQ"
```

```
Query Match             Best Local Similarity   98.8%;
Matches 1974;  Conservative  0;  Mismatches  0;  Indels  2;  Gaps  2;
Score 1952;  DB 10;  Length 1981;  Pred. No. 0;
```

```
BASE COUNT     532 a   489 c   534 g   426 t
ORIGIN
```

```
Qy     1  GGCACGAGGTCGCGGTGGAGCGGAAATTTGAAGGAAACCGGACGCGGACGCGGTTCCACCAAA    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7  GGCACGAGGTCGCGGTGGAGCGGAAATTTGAAGGAAACCGGACGCGGACGCGGTTCCACCAAA    66

Qy    61  CTGTGTGTCTGTGTCTCTGGCAGCTGGGTGGGTCCCGTGCCTTGAGTCGAGCCATCATGCCTATC   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    67  CTGTGTGTCTGTGTCTCTGGCAGCTGGGTGGGTCCCGTGCCTTGAGTCGAGCCATCATGCCTATC   126

Qy   121  CTCTCTCTGTGCACTATCTCTGCTCCCGACTTCTTCGATCACACTCCCGTGACGTGGCTGCCATC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127  CTCTCTCTGTGCACTATCTCTGCTCCCGACTTCTTCGATCACACTCCCGTGACGTGGCTGCCATC   186

Qy   181  CACTGTGGGCACACTTTTCATCTGCAAATGCCTAATCCAGTGGTTTGAGACAGCACCAAGT       240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187  CACTGTGGGCACACTTTTCATCTGCAAATGCCTAATCCAGTGGTTTGAGACAGCACCAAGT       246

Qy   241  CGGACCTGCCCACAGTGTAGAAATCCAGGTTGGGCAAAAAAGACTATTATAAACAAAACTTTTC   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247  CGGACCTGCCCACAGTGTAGAAATCCAGGTTGGGCAAAAAGACTATTATAAACAAAACTTTTC    306

Qy   301  TTTGACCTCGCCCCAGGAAGAGGAGAAATTGTCTTGGATGCAGAAATTCTTAAAGAATGAAACTG   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307  TTTGACCTCGCCCCAGGAAGAGGAGAAATTGTCTTGGATGCAGAAATTCTTAAAGAATGAAACTG   366

Qy   361  GACAGCAGGTCAAAGCTCAGCTTTCCCAGAAAAGAGAACAGGGAGAACAGGCCAGGCCATT      420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367  GACACCGTCAAAGCTCAGCTTTCCCAGAAAGACAGGGAGAACAGGCCAGGCCATT          426

Qy   421  ATCGACACTCTACGGGACACCCTGGAAGAACGCAATGCTACCGTGGAGTCCCTACAGAAC      480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427  ATCGACACTCTACGGGACACCCTGGAAGAACGCAATGCTACCGTGGAGTCCCTACAGAAC      486

Qy   481  GCCTTAAACAAGGCAGAGATGCTGTGTTCACCCTGAAAAAACAGATGAAGTTCCTGGAG       540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   487  GCCTTAAACAAGGCAGAGATGCTGTGTTCACCCTGAAAAAACAGATGAAGTTCCTGGAG       546

Qy   541  CAGCGGCAGGATGAGACAAAACAAGCTCGGGAGGAGGCCCACCGACTCAAGTGCAAGATG      600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   547  CAGCGGCAGGATGAGACAAAACAAGCTCGGGAGGAGGCCCACCGACTCAAGTGCAAGATG      606

Qy   601  AAAACCATGGAGCAAAATTGAGCTCCTTACTCCCAGAGCCAGGCGTTCTGAGGTTGGAGAGGATG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   607  AAAACCATGGAGCAAAATTGAGCTCCTTACTCCCAGAGCCAGGCGTTCTGAGGTTGGAGAGGATG 666
```

```
Qy   661  ATTCGAGACATGGGTGTGGGACAGTCAGCGTGGTGGCAGTCAGCGGTGGTCGTGTTGGCGTCC    720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   667  ATTCGAGACATGGGTGTGGGACAGTCAGCGTGGTGGCAGTCAGCGGTGGTCGTGTTGGCGTCC    726

Qy   721  CTCAAGAAAGAGACTATGAGAAATCTGAAGGAAGCTCGGAAGGCCACAGGGGAACTGGCTGAC    780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   727  CTCAAGAAAGAGACTATGAGAAATCTGAAGGAAGCTCGGAAGGCCACAGGGGAACTGGCTGAC    786

Qy   781  AGGTTGAAGAAGGACATTGTGGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACACTGAGCTG   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   787  AGGTTGAAGAAGGACATTGTGGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACACTGAGCTG   846

Qy   841  GATCAGGCCAAGTTAGAACTCAGGCCAGCAGAAGGACTTACAAAAGTGCTGACCAAGGA      900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   847  GATCAGGCCAAGTTAGAACTCAGGCCAGCAGAAGGACTTACAAAAGTGCTGACCAAGGA      906

Qy   901  ATCACGAGGCCTAAGAAAGAAGTCTGATGAATGCTCCAGAGAAGTCCTTGAGCCTTGCCTCCCG   960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   907  ATCACGAGGCCTAAGAAAGAAG-CTGATGAATCCTCCAGAGAAGTCCTTGAGCCTTGCCTCCCG   965

Qy   961  -ACCAATGAGACAGGTCAGCCGCGCCTGGTGTTTTTGGAGCCCTCGTTGGGAGATGATGAA   1019
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   966  TACCAATGAGACAGGTCAGCCGCGCCTGGTGTTTTTGGAGCCCTCGTTGGGAGATGATGAA   1025

Qy  1020  CCCGAGGCTTCACCAGCCACCCTTCGTGATGATGAGATCTCAATACCACCTTTGATGT    1079
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1026  CCCGAGGCTTCACCAGCCACCCTTCGTGATGATGAGATCTCAATACCACCTTTGATGT    1085

Qy  1080  AAATACCCCTCCAACCCAGACCTCTGGCTCCCAGCATTGCTCCCCAAGAGAGCTGCCT    1139
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1086  AAATACCCCTCCAACCCAGACCTCTGGCTCCCAGCATTGCTCCCCAAGAGAGCTGCCT    1145

Qy  1140  GGAGAGGGCACGCTGCCTCCCATGCAGAAGTCTCCTCCAGAGATGTCCAGAAGTCTCCAAGCC   1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1146  GGAGAGGGCACGCTGCCTCCCATGCAGAAGTCTCCTCCAGAGATGTCCAGAAGTCTCCAAGCC   1205

Qy  1200  GGAGTCCCAGTCTCACTGGGTGGCCAGCGGACGATGTGAGGGAGGAGGAGGAACTGGC       1259
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1206  GGAGTCCCAGTCTCACTGGGTGGCCAGCGGACGATGTGAGGGAGGAGGAGGAACTGGC       1265

Qy  1260  TGGTGCCTTCCCTCTCTTCATCCGGAATGCTGTCCTGGGTCAGAAGAACAGCCCAACAGGAC    1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1266  TGGTGCCTTCCCTCTCTTCATCCGGAATGCTGTCCTGGGTCAGAAGAACAGCCCAACAGGAC    1325

Qy  1320  CACAGCAGAAATCCCGAAGCAGCACAGATGTGGTAAGGATAGGCTTTGATGGGCTTGGAG      1379
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1326  CACAGCAGAAATCCCGAAGCAGCACAGATGTGGTAAGGATAGGCTTTGATGGGCTTGGAG      1385

Qy  1380  ACGAACAAAAATTCATCCAGCCTGAGGACACAACCATTATCCGGACCAGTGCCTGTTAAGTC   1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1386  ACGAACAAAAATTCATCCAGCCTGAGGACACAACCATTATCCGGACCAGTGCCTGTTAAGTC   1445

Qy  1440  CAAGGCCAAGAGTAAACAGAAAAGTGAGAAATGAGACTGTGAGTGCCTCCCAGCCCAA     1499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1446  CAAGGCCAAGAGTAAACAGAAAAGTGAGAAATGAGACTGTGAGTGCCTCCCAGCCCAA     1505

Qy  1500  GCTGGATACCTTCTTATGTCACTGAACGGTGACCAGAGTGAGTGCAATTAGTAGTGGGCC    1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1506  GCTGGATACCTTCTTATGTCACTGAACGGTGACCAGAGTGAGTGCAATTAGTAGTGGGCC    1565

Qy  1560  AAGACCTGGCTAAACCGGAAGTGTTTTTTTGGAGAGATGGCTGCTGACCGAGTGACCAAGAGA   1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1566  AAGACCTGGCTAAACCGGAAGTGTTTTTTTGGAGAGATGGCTGCTGACCAGTGACCAAGAGA   1625

Qy  1620  GATGCCCAGAAAAACACACTTCCTGTTCACTGCGCCCTGCACCCACCACCTGGAAGCCACA    1679
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1626  GATGCCCAGAAAAACACACTTCCTGTTCACTGCGCCCTGCACCCACCACCTGGAAGCCACA    1685

Qy  1680  TGACCAGTTTACTGTTCCGATCAGCAGGGCCTACTTCCAGTTGCCAGGGTTTTTGCTTATAG   1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1686  TGACCAGTTTACTGTTCCGATCAGCAGGGCCTACTTCCAGTTGCCAGGGTTTTTGCTTATAG   1745

Qy  1740  CTACAACCAGTGTTGCTGGACTCCTTTGTTTTTTATAGACACAGGGTCACATTGACTCTA    1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         CTACAACCAGTGTTGCTGGACTCCTTTGTTTTTTATAGACACAGGGTCACATTGACTCTA    1799
```

```
Db   1746 CTACAACCAGGTGGTGGCTGGGACTCCTTTGTTTTGTTTTATAGAACAGGGTCACATTGACTCTA 1805
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1800 AGTGGATGGGAGGTGCTGGAGGATCCTATGCAGCAGCCTGGAGGACCTCCTTGAACTCCTG 1859
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1806 AGTGGATGGGAGGTGCTGGAGGATCCTATCCAGGCGTGGAGGACCTCCTTGAACTCCTG 1865
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1860 CCTGCCTCCAGCTTATTTGCTTGAAATTATGGGGTGAGGTGGGAAAGGTTGGGG 1919
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1866 CCTGCCTCCAGCTTATTTGCTTGAAATTATGGGGTGAGGTGGTGATAGGGAAAGGTTGGGG 1925
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy   1920 AAGTTTTCTGTGTGTAAAATAAAAAAGGGATCTTTTCTTCAAAAAAAAAAAAAA 1975
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1926 AAGTTTTCTGTGTGTAAAATAAAAAGGGATCTTTTCTTCAAAAAAAAAAAAAAAAA 1981
```

RESULT 2
BC017374
LOCUS   BC017374   2024 bp   mRNA   linear   ROD 14-NOV-2001
DEFINITION   Mus musculus, TRAF-interacting protein, mRNA, complete cds.
                IMAGE:3990371, mRNA, complete cds.
ACCESSION   BC017374
VERSION   BC017374.1 GI:1692420B
KEYWORDS   MGC.
SOURCE   house mouse.
  ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2024)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-NOV-2001) National Institutes of Health, Mammalian
                Gene Collection (MGC), Cancer Genomics Office, National Cancer
                Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Gilbert Smith, Ph.D.
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Institute for Systems Biology
                http://www.systemsbiology.org
                contact: amadan@systemsbiology.org
                Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
                Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

                Clone distribution: MGC clone distribution information can be found
                through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                Series: IRAK Plate: 18 Row: d Column: 5
                This clone was selected for full length sequencing because it
                passed the following selection criteria: matched mRNA gi: 6755868.
FEATURES   Location/Qualifiers
     source   1..2024
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="MGC:13919 IMAGE:3990371"
                     /tissue_type="Mammary tumor metastasized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMTV."
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /vector="pCMV-SPORT6"
                     /note="Vector: pCMV-SPORT6"
     CDS   102..1514
                     /codon_start=1
                     /db_xref="LocusID:22036"
                     /db_xref="GI:1692409"
                     /protein_id="AAH17374.1"
                     /product="TRAF-interacting protein"
                     /translation="MPIRALCTICSDFDHSRDVAAIHCGHTHFHLQCLIQWFETAPSR
                     TCPCRIQVGKKTIINKLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQA
                     IIDFLRDTLEERNATVESLQNALNKAEMLCSTLKKQMRFLEQRQDETKQAREEAHRLK
                     CKMKTMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKAT

BASE COUNT   582 a   489 c   528 g   425 t
ORIGIN

                                                          GELADRLKDLVSSRSKLKTLNTELDDQARKLELRSRAQKDLQSADEGIMSLRKKLMILQG
                                                          TLSLPPATNETVSRILVFESPAPVEMMNPRLHQPFGDEIDLNTTFDVNTPPTQTSGSQ
                                                          HCLPKKLLCLERARSPMQNVIKKVHHVSKPESGLSLGGQRCVGELEDEELAGAFFLPIRN
                                                          AVLGQKQPNRTTAESRCSTDVVRIGFDGLGGRTKFEIQPRDTTIIRPVVKSKAKSKOK
                                                          VRIKTVSSASQPKLDTFLCQ"

Query Match   95.2%; Score 1880; DB 10; Length 2024;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1935; Conservative   0; Mismatches   0; Indels   6; Gaps   4;

Qy   24 ATTTGAAGGAAACCGGAGCGGCGGCGGCCGGGTT---CCACCAAACTGTGTCGTCGTGCACTATCTG 80
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   11 ATTTGAAGGAAACCGGAGCGGCGGCGGCCGGGTT---CCACCAAACTGTGTCGTCGTGCACTATCTG 70

Qy   81 TGGTTCCCTGGGCGTGCTTGAGTCGACCCATCAATGCCTATCCTCTCTCTGTGCACTATCTG 140
          ||||||||||||||||||||||||||||||| | |||||||||||||||||||||||||||
Db   71 TGGTTCCCTGGGCGTGCTTGAGTCGACCCATCAATGCCTATCCGCGCTCGTGCACTATCTG 130

Qy   141 CTCCCACTTCTTTCGATCACTCCCGTGACGTGGCGTGGCTGCCATCCACTCGTGGCACACATTTTCA 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 CTCCCGACTTCTTTCGATCACTCCCGTGACGTGGCGTGGCTGCCATCCACTCGTGGCACACATTTTCA 190

Qy   201 TCTGCAATGCCTAATCCAGTGGTTTGAGACGAACGTCGGAGCACCAAGTCGGAGCAGCAGTGTAG 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 TCTGCAATGCCTAATCCAGTGGTTTGAGACGAACGTCGGAGCACCAAGTCGGAGCAGCAGTGTAG 250

Qy   261 AATCCAGGTTGGCAAAAAGACTATTATTATAAACAAAAACTTTTTCTTTGTGACCTCGCCCAGGAAGA 320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 AATCCAGGTTGGCAAAAAGACTATTATTATAAACAAAAACTTTTTCTTTGTGACCTCGCCCAGGAAGA 310

Qy   321 GGAGAATGTCTTGGATGCAGAATTCTTAAAGAATAAACTGGACACGGTCACACACTTTTCA 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 GGAGAATGTCTTGGATGCAGAATTCTTAAAGAATGAACTGGACACGGTCAAAGCTCAGCT 370

Qy   381 TTCCCAGAAAAGACAGGGAGAAACGGGACAGCCAGGCCCCATTATCGACACTCTACGGGACAC 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 TTCCCAGAAAAGACAGGGAGAAACGGGACAGCCAGGCCCCATTATCGACACTCTACGGGACAC 430

Qy   441 CCTGGAAGAAGCAATGCTACCGTGGAGTGGAGTGTCCCTACAGAGAACGGCCTTAAACAGGCAGAGAT 500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   431 CCTGGAAGAAGCAATGCTACTGTGGAGTGGAGTGTCCCTACAGAGAACGGCCTTAAACAGGCAGAGAT 490

Qy   501 GCTGTGTGTTCCACCCTGAAAAAAAACAGATGAAGTTCCTGGAGCAGCCGGCGGAGTGAGACCAA 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   491 GCTGTGTGTTCCACCCTGAAAAAAAACAGATGAAGTTCCTGGAGCAGCCGGCGGAGTGAGACCAA 550

Qy   561 ACAAGCTCGGGAGGAGGAGCCCACCGACTCAAGTGCCAAGATGAAAAACCATGGAGCAAAATTGA 620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   551 ACAAGCTCGGGAGGAGGAGCCCACCGACTCAAGTGCCAAGATGAAAAACCATGGAGCAAAATTGA 610

Qy   621 GCTTCCTACTCCAGAGCGGTCTTCTGAGGTGGAGAGATGAATTCGAGACACATGGGTCGTGGGG 680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 GCTTCCTACTTCCAGAGCGGTCTTCTGAGGTGGAGAGATGAATTCGAGACACATGGGTCGTGGGG 670

Qy   681 ACAGTCAGCGCGTGGAGCAGCGTGGCGTGTCCTGCTGTCCCTCAGAAAAGAGTATGAGAA 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 ACAGTCAGCGCGTGGAGCAGCGTGGCCTGTTCGTGCGTGTCCCTCAGAAAAGAGTATGAGAA 730

Qy   741 TCTGAAGGAAGCCTCGGAAGGCCACAGGGGAAGCTTGGTCACAGTTGAAGAAGGAGATTGGT 800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731 TCTGAAGGAAGCCTCGGAAGGCCACAGGGGAAGCTTGGTCTGACAGTTGAAGAAGGAGATTGGT 790

Qy   801 GTCCCTCTAGGAGCTAAGTTGAAGACTCTTCAACACTGACCTGGATCCAGGCCAAGTTAGAACT 860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791 GTCCCTCTAGGAGTAAGTTGAAGACTCTCAACACTGACCTGGATCAGGCCAAGTTAGAACT 850

Qy   861 GAGGTCAGCCCAGAAGGACTTACAAAGTGCTGACCAGGAGAATCAGAGGCTCAAGAAGAGA 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GAGGTCAGCCCAGAAGGACTTACAAAGTGCTGACCAGGAGATCATGAGCCTCAAGAAGAA 910
```

RESULT 3

```
LOCUS       AR072729                2065 bp    DNA     linear   PAT 28-AUG-2000
DEFINITION  Sequence 1 from patent US 5948643.
ACCESSION   AR072729
VERSION     AR072729.1  GI:9999493
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2065)
  AUTHORS   Rubinfeld,B., Polakis,P.G., Lingenfelter,C. and Vuong,T.T.
  TITLE     Modulators of BRCA1 activity
  JOURNAL   Patent: US 5948643-A 1 07-SEP-1999;
FEATURES             Location/Qualifiers
     source          1..2065
                     /organism="unknown"
BASE COUNT      561 a    526 c    561 g    417 t
ORIGIN
```

Query Match 54.5%; Score 1075.4; DB 6; Length 2065;
Best Local Similarity 75.1%; Pred. No. 3.6e-297;
Matches 1460; Conservative 0; Mismatches 426; Indels 59; Gaps 7;

```
Qy   81 TGGTTCCCTGGGCTGCTTGAGTCGAGGCCATCATGGCCTATTCCTCTCTGTGTGCACTATCTG 140
        || ||||||||||||||||||||||||| ||||||||||||| |||||||||||| ||
Db   72 TGGCTGCCTGGGCCCCCTTTGAGTCCAGCCCATCATGCCTATTCCGTTGCCACTATCTG 131

Qy  141 CTCCGACTTCTTCGATCACTCCCGTGACGTGGCCTGCCAAGCTCACTGTGGCCACACTTTTCA 200
        ||||||||||||||||||||||||| || ||||||||||||||||||||||||| |||
Db  132 CTCCGACTTCTTCGATCACTCCCGACGTGGCCTGCCGCCACCACCTTCCA 191

Qy  201 TCTGCAATGCCTAAATCCAGTGGTTTGAGACAGCCACAAGTCGGACTCTACGGAGAC 260
        | |||||| ||||||||||||| ||||||||||||||||||||||||||||||
Db  192 CTTGCAGTGCCTTAATTCAGTGGTTTGAGACAGCCACCAAGTCGGACCTCTACGGAGAAG 251

Qy  261 AATCCAGGTTGCAAAAGACTATTATTAAACAAACTTTTCTTTAGACCTGCGCCGAGAGA 320
        ||||||||||| || | |||||||||||||||||||||||| ||||||||||
Db  252 AATCCAGGTTGCAAAAGAACACCATTATCAAATAAGCTCTTTCTTTGCCGCGAGGA 311

Qy  321 GGAGAATGTCTTGGATGCCAGAATTCTTAAAGAAGATGAACTCGGACAGCCGCCAAAGCTCAGGT 380
        |||||||||| || ||||||||||| ||||||||||||||||||||||||||||||
Db  312 GGAGAATGTCTTGGATGCCAGAATTCTTAAAGAATGAACTGGACAATGTCAGAGCCCAGCT 371

Qy  381 TTCCCAGAAAAAGACAGGAGAAAACGAGACCCCATTATCGACAACTCTACGGGCACAC 440
        |||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db  372 TTCCCAGAAAAGACAGGGAGAAAACGAGACAGCCGGCATCGACAACTCTCCGGGGATAC 431

Qy  441 CCTGGAAGAAACCCAATCCTACCGTGGAGTCCCTACAGAACCGCCTTAAACAAGGCACAGAGAT 500
        |||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||||
Db  432 GCTGGAAGAAACGCCAATCCTACTGTGTATTCTGCACCAGCCGCCTTGGGCAGGCACAGAT 491

Qy  501 GCTGTGTTCCACCCTGAAAAAAACAGATGAAGTTCCTGGAGCAGCAGCGGCAGGATGAGACCAA 560
        ||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||
Db  492 GCTGTGCTCCACACTGAAAAAAACCAGATGAAGTACTTAGAGCAGCAGCAGGATGAGACCAA 551

Qy  561 ACAAGTCGGAGGAGGAGCCCACCGACTCAAGTGCAAGGATGAAAAACCATGGAGACAAAATTGA 620
        |||||| ||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db  552 ACAAGCACAAGAGGAGGAGGCCCGCCTCAGGAGCAAGAACCAGAGGAGCAGAGATGA 611

Qy  621 GCTCCTACTCCAGAGCCACGGTTCTGAGGTGGAGGAGATGATGGTGGGAGACATGGGTGTGGG 680
        |||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  612 GCTTCTACTCCAGAGCACCAGCGCCCTGAGGTGGAGGAGAGATCCGAGACATGGGGG 671

Qy  681 ACAGTCAGCCGTGGGAGCAGCTGGCTGCTACTCCGCGTGTGTTACTGGTGTATCTGGTGGAA 740
        ||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||
Db  672 ACAGTCAGCCGTGGGACACGCTGGCTGTGTACTCGCTGTGTTACTGTTGTATCTGGTGGAA 731

Qy  741 TCTGAAGGAAAGCTCGGAAGGCCCACAGGGAACTGGGCTGACAGGTTGAAGAAGGATTTGGT 800
        ||| || ||||||||||| |||||||||| ||||||||||| |||||||| ||||| |
Db  732 TCTAAAAGAGGAAGCGCACGGAAGGCCCTCAGGGAGGTGGCTGACAAGCTGAGGAAGGATTTGTT 791
```

```
Qy  921 GTCTCGATGATCCTCAGGGAGAACCTTGAGCCTTGCCCTCC-GCGACCAATGAGACGGTCAGCC 979
        ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db  911 G-CTGATGATCCTCAGGGAACCTTGAGACCTCCGGCGACCAATGAGACGGTCAGCC 969

Qy  980 GCCTGGTGTTTTTGAGACCCCACCCCCTGTGTGGAGATGATGAACCTGCCTTCACCAGCCAC 1039
        ||||||||||||||| |||||||||||||||||||||||| |||||||||||||||
Db  970 GCCTGGTGTTTTTGAGACCCCACCCCCTGTGTGGAGATGATGAACCCGCCTTCACCAGCCAC 1029

Qy 1040 CCTTCGGTGATGAGAGATTGATCTCAATATACCACCTTTGATGTAAATACCCTCCAACCCAGA 1099
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1030 CCTTCGGTGATGAGAGATTGATCTCAATATACCACCTTTGATGTAAATACCCTCCAACCCAGA 1089

Qy 1100 CCTCTGGCTCCCCAGCAATTGCCTCTCCCCAAGGAAGCTGTGCCTCCTGGAGAGGGCAGCGCTCTCCCA 1159
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1090 CCTCTGGCTCCCCAGCAATTGCCTCTCCCCAAGGAAGCTGTGCCTCCTGGAGAGGGCAGCGCTCTCCCA 1149

Qy 1160 TGCAGAATGTCCTCAAGAAAGTCTCCAAAAGTCCGAGCTCTCCAGCTCTCACTGG 1219
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1150 TGCAGAATGTCCTCAAGAAAGTCTCCAAAAGTCCGAGCTCTCCAGCTCTCACTGG 1209

Qy 1220 GTGGCCAGCCGATGTGTTGTAGGAGGACTAGAATGAGGAACTGGCTGGTGGTGCCTTCCCTCTCTTCA 1279
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1210 CTAGGGACACAACCATTATCCGACCAGTGCTGTTAAGTCCAAGGTGCCCTCTGTTCA 1269

Qy 1280 TCCGGAATGCTGTCTGGGTGCAGAAAACAGCCCAACAGCCCAGAAGTCCCGTTGCA 1339
        ||||||||||||||||||||| || |||||| |||||| |||||||||
Db 1270 AAGTGAGAATAAAAGACTGTGAGTTCTGCCTCCCAGCCCCAAGCTGGATAACCTTTTTATGTC 1329

Qy 1340 GCACAGATGTGGTAAGAATAGGCTTTGATGGGCTTGGAGGACGACAGAAATTCATCCAGC 1399
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1330 GCACAGATGTGGTAAGAATAGGCTTTGATGGGCTTGGAGGACGACAGAAATTCATCCAGC 1389

Qy 1400 CTAGGGACACAACCATTATCCGACCAGTGCCTGTTAAGTCCAAGGTGCCAAGAGTAAAACAGA 1459
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1390 AGTGAACGGTGACCAGAGTCATGTTTGCAATTAGTGGGGCCAAGAGACCTGGCTAACCGGAAG 1449

Qy 1460 AAGTGAGAATAAAGACTGTGAGTTCTGCCTCCCAGCCCTGGATAACCTTCTATGTC 1519
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1450 AAGTGAGAATAAAGACTGTGAGTTCTGCCTCCCAGCCCTGGATAACCTTCTATGTC 1509

Qy 1520 AGTGAACGGTGACCAGAGTCATGTTTGCAATTAGTGGGGCCAAGAGACCTGGCTAACCGGAAG 1579
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1510 AGTGAACGGTGACCAGAGTCATGTTTGCAATTAGTGGGGCCAAGAGACCTGGCTAACCGGAAG 1569

Qy 1580 TGTTTTTTGGAAGATGGCTCCTCCTTGGACCAGTGCCCAGAAAAACACACTT 1639
        ||||||||||||||||||||||||||||||||||||||||||||
Db 1570 TGTTTTTGGAAGATGGCTCCTCCTTGGACCAGTGCCCAGAAAAACACACTT 1629

Qy 1640 CCTGTCGTTCACTGCGGCCCTCCACCACACTGGGAAGGCCACATGACAGATTGA 1699
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1630 CCTGTCGTTCACTGCGGCCCTCCACCACACTGGGAAGGCCACATTACTGTTCCGA 1689

Qy 1700 TCAGCAGGGCCTACTTCCAGTTGCAGGGTTTTGCTTGCTTCTTTATATAGCTACAACCAGGTGTGGCTGG 1759
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1690 TCAGGAGGGCCTACTTCCAGTTGCAGGGTTTTGCTTATAGCTACAACCAGGTGTGGCTGG 1749

Qy 1760 ACTCCTTTTGTCTTTTTATAGAACAGGGTCACATTGACTCTAAGTGAGTGCTGGGAG 1819
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 1750 ACTCCTTTTGTTTTTTATAGAACAGGGGTCACATTGACTGGAGTGCTGGGAG 1809

Qy 1820 GATCCTATGCAGGCTGGAGACCCTGCGCTTGAACTCCTGCCTCCCAGCTTATTGCT 1879
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 1810 GATCCTATGCAGGCTGTAGGCTGTGAGGCTGGAGACCCTGCTGGAACTCCTGCCTCCCAGCTTATTGCT 1869

Qy 1880 TGAAATTATGGGGTGAGGTGGTGATAGGGAGAGGTTTGGGGAAAGGTTTTTCTGTGTAAAATAA 1939
        ||||||||||||||||||||||||||||||||||||||| || |||||||||||||||||
Db 1870 TGAAATTATGGGGTGAGGTGGTGATAGGGAGGTTGGGGGAAGTTTTTCTGTGTAAAAT-A 1928

Qy 1940 AAAGGGATCTTTTTCTTCTTCTAAAAAAAAAAAAAAAAAAAAAAAAAA 1975
        ||||||||||||||||||||||||||||||||||||||||||
Db 1929 AAAGGCATCTTTTTCTTCTTCTAAAAAAAAAAAAAAAA 1964
```

```
Qy    801 GTCCTCTAGGAGCAGCAAGTTGAAGACTCTCAACACTGAGCTGGATCAGGCAGCCAAGTTAGAAACT  860
          |||||||  ||| |||||| || |||||||||||||  | |||| |||||||||||||||||||||||
Db    792 TTCCTCCCAGAAGCAAAGTTGCAGAACGAGTCTACTCTGAATTGGATCAGGCAGCCAAGTTAGAAACT  851

Qy    861 GAGGTCAGCCCAGAAGGAGACTTACAAAAGTGCTGAGACGAGATCACGAGCCTAAGAAAGAA        920
          | |||||||||||||| ||| |||||||||||||||||  ||||||||||||||| ||||||||
Db    852 GAAGTCAGCCCAGAAGGAGACTTACAGAGTGCTGACAAGGAAAATCATGAGCCTGAAAAAAGA      911

Qy    921 GTCTTGATGATCTCCAGCATTGAGCCTGCTCC--GGACCAATGAGACGGTCAGCC                979
          ||||  | ||| ||||||||||||||  |||||  |||  |||| ||| ||||||||||
Db    912 G-CTAACGAATGCTGCAGGAAAACCTTGAACCTGCACCAGTGTGAGAACTGTCGACC          970

Qy    980 GCCTGGTTTTTGAGAGCCCAGCCCCCTGTCGAGAGAATGGGAACCCAGGCCTTCACCAGCCAC    1039
          |||||||||| ||||| ||| |||| |||||  |||||||  ||| ||| |||||| |||||
Db    971 GCCTGGTTTTTAGAGAGCCCCAGCCCCTGTGGA---GGTGAATCTGAAGCTCCGCCGCCCAT    1027

Qy   1040 CCTTCGGTGATGAGATTGATCTCAATACCACCTTTGAATGTGTAAAATACCCCTTCCAACCCAGA    1099
          |||||||||| || ||| ||||||||||||||||| ||||||| |||| ||| |||||||||||
Db   1028 CCTTCCGTGATGAATATTGATCTCAATGCTCATTCTACCTTTGATGTGGGATACTCCCCAGCCGGC    1087

Qy   1100 CCCTGGCTCTCCCAGCATTGCCTCCCCAAGAAGCTGTGCCTGGAGGAGGCACGCTCTCCA      1159
          |||| |  |||||||||| ||  | |||||  | ||||||| ||| || || |||| |||
Db   1088 CCTCCAGCTCCCAGCCATTGGTTACTAGGCAGAAAGTCTCCACTCACTCGGTCCACTCCCAA    1147

Qy   1160 TGCAGAATGTCCTCAAGAGAGGTGCACAAAGTCTTCCAACGTCCCACCTCCTCACTGG      1219
          |||| |||  | ||||| |||||  |||| |||| ||||||| ||||| ||||||||
Db   1148 TTCAGGAGTGTCCCAAGAAGAGAATGCAAAGGCCCCACGTCCCCAGCTCTCACTGG         1207

Qy   1220 GTGGCCAGGCGAATGTGTAGAGGAAGCTGTCCCTGGAGGAGGCAGCGCTCTCCA          1279
          ||||||||||| || ||||||| |||||||||| ||| |||||||| ||| |||
Db   1208 GTGGCCGAGCGTGTTGCAGAGCAGTGGTTCCGCCTTCCCTCCCTCTCTTTG           1267

Qy   1280 TCCGGAATGCTGTCCTGGGTCAGAAAAACAGGCACACCCAAGTCCCACCTCTCACTGG    1339
          |||| ||| | ||  ||| | ||  |||||||||||| | |||||||| || ||||||
Db   1268 TCCGGAATGCCATCCTAGGCACCCAGAAAACAGCCTCAGAGTCTGGCATCCTCTTTGCA    1327

Qy   1340 GCACAGATGTGTGGTAAGAATAGGCCTTTGATGGGCTTTGGACGGCAACACAAAATTCATCCAGC    1399
          |  | ||||||  |||| ||| ||| || ||||| ||| ||| ||||||| |||| |||||||
Db   1328 CTACTGACACAGTCATGACCGCCCATTGCCTGTTCCAAGACCAACAGGTTAAGCACAGA        1387

Qy   1400 CTAGGGACACAACCATTATCCGAACCAGTGCCTGTTAAGTCCAAGCAGTCAAAACAGA      1459
          | |  || ||  |||| |||| ||  || ||| |||||  |||| |||| |||||||
Db   1388 CTACTGACACATGATCGGCCCATTGCCTGTTAAGTCCAAGCAAGGGTTAAGCAGA          1447

Qy   1460 AAGTGAGAATAAAAGACTGTGAGTTCTGCCTCCCAGCCCAAGCTGGATACCTTCTTTATGTC    1519
          || | ||  | |||   | |||||||| ||||||| |||||  | ||||||| ||||| ||
Db   1448 GGGTGAGGTGAAGAACAGCCTTCTCTTCCAGCCAGAGTACACCTTCCGTGGT             1507

Qy   1520 AG-----------------TGAACGGTGACAGAGTGAGTGTTGCAATTAGTGGGGCAAGAC    1564
          ||                 |  | |||||| |  | | ||| |||||| | |||||||||
Db   1508 CGTGAGAACAAGTGAGTCTGACCAATGGCAACACACACGTTGTAGGTCAAGGA             1567

Qy   1565 CTGGCTAACCGGAAGCTGTTTTTTGGAAGAATGGCTCCTTCTTGGACC----------    1608
          |||||  || |||| ||| |||||||| |||| || || ||||||||
Db   1568 CTGTTCCAGCCAGGGCGTTTTCGGACAGAGAGCCCCACTTCCCTGTGCGAGGTGTAAG    1627

Qy   1609 --------AGTCCAAGAGAGATGCCCAGAAAAAACAACTTCCTGTTTGTGT          1652
                  ||  || | | || |||  |||  |  | |||| | |||| ||
Db   1628 GGCAGGACAAAACAGTGAGGGTGTGACAGCCCCACCCAGAGACTGCGGGGCC            1687

Qy   1653 CGCCCTGCACC--ACACTGGGAGGACCACCAGTTTACTGTGTTCGATCAGCAGGGCC      1710
          |||| | |||   ||  |||| ||| || |||| |||| | |||| |||||||||||
Db   1688 TGCCCCACTCCTACGGACTGGGACTGGAACTGACATGACACCGACCACCAGCCTTCCA    1747

Qy   1711 TACTTCCAGTTGCAGGGTTTTGCTTATAGCTACAACACCAGGTGTGGCTGGACTCCTTCTTTGT    1770
          ||| || | ||| ||  | |||| ||| || ||| ||||| |||||| |||||| ||| ||||
Db   1748 TGCTCCTGTTGCCACGGCTCTGTTTTTATATAGGCCATGCTGGTGTCAGGACTCTTTTCTGG    1807

Qy   1771 TTTTATAGAACAGGGGTTCACAATGGACTGCTGGAGGAGTGCTGGGAGACTCCTATGCA      1830
          |||| |||  ||  |  | ||  |||| |||||||| |||| ||||| |||| || |
Db   1808 GCCTGGAGACCACGGTCACACTTGTTGACTGTCTCGTGGACCAG-------AGTGCTTGAGGC    1862

Qy   1831 GGCTGGAGGAGACCCTCGCGCTTGAAACTCCTGCCTCCAGCGCTTATTGCTTGAAAATTATGG    1890
          || ||||| || ||  | |||| ||| ||| || | ||| ||||| |||| ||| ||| |
Db   1831 GGAGAATGTCTTGGATGCAGAATTCTTAAAAGAATGAACTGGGACAAATGTCAGAGCCCAGCT    1890
```

RESULT   4
BD006990

LOCUS   BD006990   2065 bp   DNA   linear   PAT 31-JAN-2002
DEFINITION Modulators of BRCA1 activity.
ACCESSION   BD006990
VERSION   BD006990.1 GI:18635361
KEYWORDS   JP 2001502893-A/1.
SOURCE   unidentified.
  ORGANISM unidentified
            unclassified.
REFERENCE   1 (bases 1 to 2065)
  AUTHORS   Lubinfeld,B., Porakisu,P., Rigenferuta,C. and Buon,T.T.
  TITLE   Modulators of BRCA1 activity
  JOURNAL   Patent: JP 2001502893-A 1 06-MAR-2001;
            ONIX PHARMACEUTICALS INC
COMMENT   Unidentified
OS   Unidentified
PN   JP 2001502893-A/1
PD   06-MAR-2001
PF   06-AUG-1997 JP 199512659
PR   04-SEP-1996 US 60/025601
PI   BONY LUBINFELD,PAUL PORAKISU,CAROL RIGENFERUTA,TERIRIN T BUON
PC   C12N15/09,A61K38/00,C07K14/435,C07K16/18,C12N1/15,
PC   C12N1/19,
PC   C12N1/21,C12N5/10,C12P21/02//C12P21/08,C12N15/00,C12N5/00, PC
            A61K37/02
CC   Strandedness: Double;
CC   Topology: Linear;

FEATURES   Location/Qualifiers
     source   1..2065
                     /organism="unidentified"
                     /db_xref="taxon:32644"
FH   Key   Location/Qualifiers
FT   CDS   103..1512.
                     1..1512,

BASE COUNT   561 a   526 c   561 g   417 t
ORIGIN

Query Match   54.58; Score 1075.4; DB 6; Length 2065;
Best Local Similarity 75.1%;
Matches 1460; Conservative 0; Mismatches 426; Indels 59; Gaps 7;
                     Pred. No. 3.6e-297;

```
Qy    81 TGGTTCCCTGGGCTGCTTGAGTCGGAGCCATCATGCCTANCCTCTCTCTGTTGACTATCTG    140
         |||  ||| ||||||| || |||||||||| ||||||| ||| |||| |||||||||||
Db     1 TGGCTGGCTGGGCCCCCTTCGGGACGCCATCATGCCTATCCGTGGTCGGCTCTGCCACTATCTG    131

Qy   141 CTCCGACTTCTCGATCACTCCCGTGACGTGGCTGCTCCATCACTGTGGCCACACTTTTCA    200
         | ||||||| ||||||||||| | |||| || ||| |||||||||||||||| ||  ||
Db   132 CTCCGACTTCTCGATCACTCCCGACGTGGCGCGCCACCACTGTGGCCACACCTTCA      191

Qy   201 TCTGCAATGCCTAATCCAGTGGTTTGAGACAGCACCACCACAGTGTGCCCACCGTGTAG    260
         ||  | | |||||| |||| |||| || ||| |||||| | |  ||  |||||||| |
Db   192 CTTGCAGTGCCTAATTCAGTGGTTTGAGACACACCACCAAGTCGGACCCGCCGCAGTGCCG    251

Qy   261 AATCCAGGTTGGCAAAAGACTATTATAAAACAAAACTTTTCTTTGACCTGCCCCAGGAAGA    320
         |||||||| |||||||| || | | |||| ||||  | |||  | |||| |||||||||
Db   252 AATCCAGGTTGGCAAAAGAACCATTATCAATAAGCTCTTCTTTGACCTGCTGCCCAGGAGGA    311

Qy   321 GGAGAATGTCTTGGATGCAGAGAATTCTTAAAAGAATGAACTGGGACAAATGTCAGAGCCCAGCT    380
         |||||||||| ||||| ||||| |||| ||||||| || |||| ||| |||||||||||||| |
Db   312 GGAGAATGTCTTGGATGCAGAATTCTTAAAAGAATGAACTGGGACAAATGTCAGAGCCCAGCT    371
```

```
Qy    381  TTCCGAGAAAGACAGGGAGAAACGGGGACAGGGCCCCATTATCGACACTCTTACGGGACAC  440
           ||||||   |||||||||||||||| ||||||||||||||||||||||||| |||||||||
Db    372  TTCCCAGAAAGACAGAGAGAAACGGAGACAGCAGCCAGGTCATCAATCGACACTCTCGGGATAC  431

Qy    441  CCTGGAAGAACGCAAATGCTACCGGTGGAGTCCCTACAGAAACAAAGGCAGAGAT  500
           ||||||||||||||||||||| |||||||  ||||||||||||||||||||||||
Db    432  GCTGGAAGAACGCAAATGCTACTGTGTATCTCTGCAGCAGCCGGCAAGGCCGGAT  491

Qy    501  GCTGTGTTCCACCTTGAAAAACAGATGAAGTTCCTGGAGCGGCCAGGGGATGAGAGGACCAA  560
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    492  GCTGTGCTCCACACTGAAAAAGCACAGTGAAGTACTTAGAGCAGCACGAGGCAGATGAGAGACCAA  551

Qy    561  ACAAGCTCGGGAGGGAGGGCCACCGAGTCACAAGTGAGAAGTGACAAATTGA  620
           |||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    552  ACAAGACAAGAGAGGCGGGCCCCCGGGTCAGGAGCAGATGAAGAACCATGGAGCACAGGATTGA  611

Qy    621  GCTCCTACTTCCAGGACCCAGCGGTTCTTGAGGTTGGAGAGATGGTGTGGG  680
           |||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    612  GCTTCTTACTTCCAGGACCCAGCCCCTGAGGGTGGAGGAGAGATCCGAGACACATGGGGTGTGGG  671

Qy    681  ACAGTCAGACGGGTGGGACGCAGCTGGTCTGTGTACTGCGTGTGTCCTCAAGAAGAGTATGAGAA  740
           |||||||||||||||||||||||||||||||||  || |||||||||||||||||||||||||||
Db    672  ACAGTCAGCGGTGGAACAGCTGGCGTGTGTACTGTGTGTCTCTCAAGAAAAGAGTACCAGAA  731

Qy    741  TCTGAAGGAAGCTCGGGAGGGCCACACGGGGGCCCACAGGGTTGAAGAGGGTTGGT  800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732  TCTAAAAGAGGGCACGGGAAGGGCTCAGGGAGGTGGCTGACAAGCTGAGGGAAGGATTTGTT  791

Qy    801  GTCCTCTAGGAGCAAGTTGAAGAGCTCTCAACACTGGAGCTCTGAAGTTAGAAGT  860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    792  TTCCTCCAGAAGCAAGGAAGTTGCAGACAAGTTCGAGACACATCAGGGCGAAGTTAGAACT  851

Qy    861  GAGGTCAGCCCAGAAGGGACTTACAAAAGTGCTGCACCAGGAGAATCACCGAGCCTAAGAAGAA  920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    852  GAAGTCAGCGCCAGAAGGGACTTACAGAGTGGGCTGTGTACTGGTGTCTCTCAAGAAAGAGTT  911

Qy    921  GTCTCGATGATCCTCCAGGGGAAACTTGAGCCTGCCTCC-GGGACCAATGAGACGGTCAGCC  979
           ||| ||| |||||||||||||||||||||||||||||||  |||| ||||||||||||||||
Db    912  G-CTAACGATCGTGCAGGGAAAACCTTGAAACCTGCCCAGTGACAGTGAGGACTGTCGACC  970

Qy    980  GCCTCGTTTTTGAGAGGCCCAGCCCCTGTTGGAGATGAACCGGAGGCTTCACCACCAC  1039
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    971  GCCTGGTTTTGAGACCCCAGCCCCCTGTGGA---GGTGAATCTGAAGCTCCGCCGCCAT  1027

Qy   1040  CCTTCCGTCGGTGATGAGATTGATCTCAATACCACTTTGATGATGTAAAATACCCTCCAACCAGA  1099
           |||| || |||||||||||||||||||||||||||||||||||  |||| ||| || |
Db   1028  CCTTCCGTCGATGAATGTGATGGAGGCTCGGCCACGGTCACTTGTCTGTCTGTGGAACTCCCCACCCGGC  1087

Qy   1100  CCTCTCTGGCCTCCCCAGCTGTTGCTCTCCCCAAGAGCTCTGGACGGCACGCGTCTCCCA  1159
           ||| |||||||||||||||||||||| || |||||||||||| |||||||||||| ||||
Db   1088  CCCTCCAGCTCCCAGCCCATGGTTACTACGAAAAAACTTTGCCTAGAGAAGTCACACTCCCAA  1147

Qy   1160  TGCACGAAATGTCCTCAAGAAGGTGCACAAGTACTCCAAGCGGGAGGTCCCGGACTCCACTGG  1219
           ||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1148  TTCAGGATGTCCCAAGAAGAGATATCAAAAGCCGAGGAGCGTCCCAGCCTCTCACTTCA  1207

Qy   1220  GTGGCCAGCGGATGTGTGTAGGAGATTGAGGAGGAACTGGCTGGTGCTTCCTCTCTTCA  1279
           ||||||||||| |||||||||||||||||||||||||||| |||||| || ||| || |
Db   1208  GTGGCCGAGCTGTCCAGGAGGAGCCAGATGAGGAACTGGTTGTGGTTGGCCTTCCCTATTTTG  1267

Qy   1280  TCCGGAATGCTGCTTGTCTTGGGTTCAGAAAACAGGCCACACAGAGAATCCCGAAGCA  1339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1268  TCCGGAATGCCATCCTAGGCCAGAAAACAGCCCAAGGATGAGTCCTCCTTGCA  1327

Qy   1340  GCACAGAGTGTGGTAAGGAATAGGCTTTGAGGGCTTGGAGGACGAACAAAAATTCATCCAGC  1399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1328  GCAAAGATGTGGTAAGGACAGGCTTCGATGGGTCGGTGGCGGACACAAAAATTCATCCAGC  1387

Qy   1400  CTAGGGACACAACCATTATCCGACCAGTGCCTGCCTGTTAAGTCACAGAGTAAACAGA  1459
           ||| |||| || |||||| |||||| |||||||||||||||| ||| |||| || |||
Db   1388  CTACTGACACAGTCATGATCCGCCCATTGCCTTGTTAAGCCCAAGAGGTTAAGCAGA  1447

Qy   1460  AAGTGAGAATAAAAGACTGTGAGTGTTCTGCCTCCCAGCACCCCAAGCTGGATACCTTCTTTATGTC  1519
           ||| |||| |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db                                                                      
```

```
Db   1448  GGGTGAGGGTGGAGACAGCAGTGCCTTCTCGTTCCAGGGCCAAGCTGGACACCTTTCTGGT  1507

Qy   1520  AG--------------TGAACGGTGACCACAGAGTGATGTTTGCAAATTGGTGGGCCAAGAC  1564
           ||              |||||||||||||||||||||||||||||||||||||||| |||||
Db   1508  CGTGAGAACAGTGAGTCTGACCAAATGGCCAGACACGACCTGCAAACTTGTAGGTGCAAGGA  1567

Qy   1565  CTGGCTAACCGGAAGTGTGTTTTGGAAGATGGGCTTCCTCTGGACC-----------  1608
           || || ||||||||||||||||||||||||||||||||||||||||           
Db   1568  CTGTCAGGCACGGGGTTTTGTGGACAGAGAGCCCCACTTTCGGACACCCAGGTGTAAG  1627

Qy                   AGTCCAAGAGAGAATGCCCAGAGAAAACACACTTCTGTTCACTG  1652
                        |||| |||||||||| |||||||||||||||||||||||||||
Db   1628  GGCAGAGAAAACAGGTGAGGGTGTGACACCCAGAGGACTGCTTCCCTGCCCTCACCC  1687

Qy   1653  CGCCCTCCACC--ACACTGGGAAGCACCACATGACCAGTTTACTGTTCCGATCAGCACCGGC  1710
           |||||||||||  ||||| ||  |||||||||||||||||||| |||||  |||||||||||
Db   1688  TGCCCCACTCCTTACGACTGGGGAGCTGACAATGACCAGCCACCGCACTGGAGAGGTCC  1747

Qy   1711  TACTTCCAGTTGCAGGTGTTTTCCTTATAGCTACAACCAGGTGTGGCTCCTTTTGT  1770
           ||| ||||||||||| |||||||||||  |||||||||||||||||||||||||||
Db   1748  TGCTCCTGTTGCCCAGGCCTCCGTTTTATGAGCCATGAATGACATGAAGGGGTAGCTT  1807

Qy   1771  TTTTATAGAACACAGGGTCACATTGACTCTAAGTGACGTGGGAGGTGGCTGGGGATCATTCA  1830
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1808  GCCTGGAGACCACGGTCACTTGTTGACTGGCTCTGTGGACCAG-----AGTGCTTGAGGC  1862

Qy   1831  GGCTGGAGGACCCCTGGCGTTGAAACTCCTCGCCTCCGCCGTTATTGCTTAAAATTATGG  1890
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   1863  ATCTCAGGCAGCCTCAGCCCAAGCTTCTACCTTGACCTTTGACTTGCTTCTCAGGCATAGCCT  1922

Qy   1891  GGTGAGGTGGTCGATAGGGAAAAGGTTGGGGGGAAGTTTTCTGGTGTAAAATAAAGGGATCTT  1950
           ||                  |||||||||| |||||||||||||| |||||||||||||||||
Db   1923  GGGCCAAGCACAGGGTGGGGAATGTGAGGATAAGCATGGGAGGAGGATGGGGAAGGAGATTT  1982

Qy   1951  TTCTTCAAAAAAAAAAAAAAAAAAAAAAAAA  1975
           ||||||||||||||||||||||||||||||||
Db   1983  TCATGTAAAAATAAAAAATTAAAAAAAAA  2007
```

```
RESULT   5
BC019283
LOCUS       BC019283               2002 bp    mRNA    linear   PRI 19-DEC-2001
DEFINITION  Homo sapiens, TRAF interacting protein, clone MGC:3999
            IMAGE:2821007, mRNA, complete cds.
ACCESSION   BC019283
VERSION     BC019283.1  GI:17939476
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2002)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL:: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
```

```
           Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
           Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
           George Yang, Scott Zuyderduyn, Marco Marra,

           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAL Plate: 9 Row: 1 Column: 1
           This clone was selected for full length sequencing because it
           passed the following selection criteria: matched mRNA gi: 5032194.

FEATURES             Location/Qualifiers
     source          1..2002
                     /organism="Homo sapiens"
                     /db_xref="LocusID:10293"
                     /db_xref="taxon:9606"
                     /clone="MGC:3999 IMAGE:2821007"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             88..1497
                     /codon_start=1
                     /product="TRAF interacting protein"
                     /protein_id="AAH19283.1"
                     /db_xref="GI:1793477"
                     /translation="MPIRALCTICSDFDHSRDVAAIHCGHTFHLQCLIQWFETAPSR
                     TCPQCRIQVGKRIIINKLFFDLAQEEENVLDAEFLKNELDNVRAQLSQKDKEKRDSQV
                     IIDTLRDTLEERNATVVSLQQALGKAEMLCSTLKKQMKYLEQODETKQAQEEARRLR
                     SKMKTMEQIELLLQSQRPEVEMIRDMGVQGSAVEQLLAYVCVSLKKYEVENLKEARKAS
                     GEVADKLRKDLFSSRSKLQTVVSELDQAKLELKSAQKDLQSADKEIMSLKKLTMLQE
                     TLNLPPVASETVDRLVLESPAPVEVNLKLRRPSFRDDIDLNATFDVDTPPARPSSSQH
                     GYEKLCLEKSHSPIQDVPKKICKGPRKESQLSLGGQSCAGEPDEELVGAFPIEVRNA
                     ILGQKQPKRPRSESSSCSKDVVRTGFDGLGGRTKFIQPTDTVMIRPLPVKPKTKVKQRV
                     RVKTVPSLFQAKLDTFLWS"

BASE COUNT      516 a     517 c     556 g     413 t
ORIGIN

        Query Match        54.3%;  Score 1072.2;  DB 9;  Length 2002;
        Best Local Similarity   75.0%;  Pred. No. 3e-296;
        Matches 1458;  Conservative   0;  Mismatches   428;  Indels   59;  Gaps   7;
```

```
Qy     81 TGGTTCCCTGGGCTGCGTTGAGTCGGACGCCATGGCCATCCTCCTCTCTGTGCACTATCTG 140
          ||| | |||||| | ||  ||||||| |||||||||||||||||||||||||||||||||
Db     57 TGGCCTGGTCGGGGCCCCTTCAGTCCAGCCATCATCGCCATCCTCTGTGCACTATCTG 116

Qy    141 CTCCGACTTCTTCGATCACTCACTCCCCGTGACCGTGCCGTCCCACACTTTTCA 200
          ||| ||||||||  |||||  ||||||| ||||||||||| || ||| ||||| ||||||
Db    117 CTCCGACTTCTTCGATCACTCCCGCCGACGGTGGCCCCCATCCACTGCCACACTTTTCA 176

Qy    201 TCTGCAATGCCTAATCCACTGGTTTGAGACAGCAGCACCAGTCCCACAGTGTAG 260
          || || |||||||||||  ||||||||| ||||| ||  ||||| || |||||||
Db    177 CTTGCAGTGCCTAATTCAGTGGTTTGAGACAGCAGCACCAGTCGGACCTGCCCACAGTGCCG 236

Qy    261 AATCCAGGTTGGCAAAAAGACTATTATTAAACAAACATTTTCTTTGACCTCCCCAGGAAGA 320
          ||||||||||| |||| |||| ||||||| |||| |||||| |||||||||||||||||
Db    237 AATCCAGGTTGGCAAAAAGAACCATTATCAATAAGCTCTTCTTTGATCTTGCCCAGGAGGA 296

Qy    321 GGAGAAATGTCTTGGATGCAGAAATTCTTAAAGAATGAAACTGGACACGCGTCAAAAGTCTCAGCT 380
          ||||||||| ||||| ||||||||  |||||||||| |||||||| || ||| ||||||
Db    297 GGAGAAATGTCTTGGATGCAGAAATTCTTAAAGAATGACAATGCAGTGAGAGCCCAGCT 356

Qy    381 TTCCCAGAAAGACACAGGGAGAAAACGGGACAGCAGCCATTATCGACACTCTACCGGGACAC 440
          ||||||||| |  ||||| |||| ||  |||| |||| ||||  ||||| ||||||||
Db    357 TTCCCAGAAAAGAAGAAGGACAGAACGGCCCAGGTCATCATCGACACTCTGCGGGATAC 416

Qy    441 CCTGGAAGAACGCAATGTCACCGTGGAGTCCCTACAGAACGCCTTAAACAAGGCAGAGAT 500
          |||||||| ||||  ||||||||  || |||||| |||||| || ||| |||| ||||
Db    417 GCTGGAAGAACGCAATGCTACTGTGGTATCTCCGCAGCAGGCCCTTGGGAAGGCGGAGGA 476

Qy    501 GCTGTGCTCCACACTGAAAAAGCAGATGAAGTACTTAGAGCAGCAGGAGATGAGACCAA 560
          |||||| || || || |||||||||| ||||| |  |||||||| |||| |||| ||||
Db    477 GCTGTGCTCCACACTGAAAAAGCAGATGAAGTAGCTGGGGACAGCCCTGAGGGTGCACTG 536
```

```
Qy    561 GCTCCTTACTCCAGGCCCAGCCATCAGTGCCAAGATGAAAAACCATGGAGCAAAATTGA 620
          ||| |||| |||||||||| |||| |||||| |||||||||| ||||| |||||||| |||
Db    537 ACAAGCAACAAGAGGAGGCCCGGGCTTCAGGTGGAGGACATGGGTGTGCAGATTGA 596

Qy    621 GCTCCTTACTCCAGACCCAGCCCCTCAGGTTGGAGGAGGACATGGAGCAGCCACAT 680
          |||||||| |||||| ||| || || ||| |||| ||||||| ||||| || ||||
Db    597 GCTTCTTACTCCAGACCCAGCCCCTCAGGTTGGAGGAGGACATGGAGCAGCCACAT 656

Qy    681 ACAGTCAGCCGGTGGAGCAGCTGCTGGCTTACGTGTGCGTGTCCCTCAAGAAGTATGAGA 740
          ||||||||||| |||||||||| ||||||||||||||| |||||||||||||||| ||||
Db    657 ACAGTCAGCCGGTGGAACAGCTGGCTGTACGTGTGCGTGTCTCTCAAGAAAGTACGAGAA 716

Qy    777 TTCCTCCAGAGAAGCAAGTTGCAGACAGTACTCTGAATTGATCAGCGGCCAAGTTAGAACT 836
          |||||||||| || ||||||||| || |||| ||  ||||| |||| || ||||||||||
Db    717 TCTAAAAGAGGCACAGAAGCCCTCAGGGAGCTGGGACAAGCTGAGGAGGAGATTTGTT 776

Qy    801 GTCCCTCTAGAGGCACAAGTTGAAGACTCTCAACACTGAGCTCAGGGGAGGCAAGTTAGAACT 860
          || ||||||||| || |||| || || | |  ||||| || ||||| ||| ||||||||
Db    777 TTCCTGCAGAAGCAAGTTGCAGACAGTACTCTGAATTGATCAGCGGCCAAGTTAGAACT 836

Qy    861 GAGGTCAGCCCAGAAGGACAGTTACAAAGTGCTGACCAGAGAGATCACGAGCCTAAGAAA 920
          || |||||||||||||| |||||| || |||||| ||||||| ||| ||| ||| || ||
Db    837 GAAGTCAGCCCAGAAGGACTTACAGAGTGCTGACAAGGAAATCATGAGCCTGAAAAAGAA 896

Qy    921 GTCTGATGATCCTCCAGGGGAACCTGAGCCTGCCTCC-GCGACCAATGAAGAGACGGTCAGCC 979
          || |||| | |||||||| |||| || || || ||| |||| || || || ||||||
Db    897 G-CTAACGATGGCTGCAGGAAACCTTGAACCTGCCCACGTGACTCAGGGCTGTGTGACC 955

Qy    980 GCCTGGTTTTTTGAGGACGCCCAGCCCCTGTGGGATGATGAAGACCCGAGCGGCTTCACCGACC 1039
          ||||||||| ||||| || |||||||||  |  || || ||| || |||  || ||||
Db    956 GCCTGGTTTTTGAGGACGCCCAGCCCCTGTGGA---GGTGAAATCGGAAGCTCGCCGGCCAT 1012

Qy   1040 CCTTCGGTGATGAGATTGATCTCAATACCACCTTTGATGTAAATACCCCTCCAACCCAGA 1099
          ||||||| | || ||| ||  || |||||| |||| || | ||  |||| ||||||||
Db   1013 CCTTCGGTGATGATATTGATCTCAATGGCACTACTCCCCAGCGCCCG 1072
```

```
Qy   1100 CCTCTCCTGGTCCCCAGCCATTGCCTCCCCAGGAAGCTGTGCCCTGGAGGACGAGCCTCTCCA 1159
          |||| ||| ||||||||| ||  ||  ||| ||||| |||| ||||||| ||  |  ||||
Db   1073 CCTCCAACTCCCCAGCCATGTTGTACTAGGAGAGCCAGATGAGGAGCTGGTCGGAGCCTTCCCA 1132

Qy   1160 TGCAGAAATGTCCTCAAGGAGAAGGTGCACAAAGTCTCCAGGAGTCCAGGGTCCAGCTGG 1219
          ||||| ||||| |||||| || || || || |||| ||  || || || ||| ||| |||
Db   1133 TTCAGAGATGTCGGGGTGCAGGGCAGTGCAGTGGAGCAGCTGCTGGCCTACGTCTGTGT 1192

Qy   1220 GTGGCCAGCGGATGTGTAGGAGAGAGCTAGATGATGAAGACCCAGGAGACAGAGTCAGC 1279
          |||| |  |  |||||| || || ||| |||| ||| || ||| ||  || ||||  |
Db   1193 GTGGCCAGCGGATGTGTAGGAGAGAGCTAGATGATGAAGACCCAGGAGACAGAGTCAGC 1252

Qy   1280 TCCGGAATGCGTGTCTCGGGTCAGAAACAGCCCAACAGGAGAAATCCCGAACAGCA 1339
          || || || ||| |||| || || || || ||| |||  ||||  ||||| || ||
Db   1253 TCCGGAATGGCCATCCTAGGCCACGAAGAGCCCAGATGAGGTCAGAGTCCTGTTTGCA 1312

Qy   1340 GCACAGATGGTGGTAAGGAATAGGCTTTGATGGGCTTTGGAGGACGAACCAGCAGC 1399
          || ||| || ||  || |||||||||| || |||  ||||||| |  |||  ||||
Db   1313 GCAAAGATGTGGTAAGGACAGGCTTCGGATGGCCGGGACAAAAATTCATCCAGC 1372

Qy   1400 CTAGGGACACAACCATTATCCGACACAGTGGGTCTGTTTAGTGACAGTAAAGTAAACAGA 1459
          || ||| ||| | ||| |  || ||||| |  || || || || ||| ||| |||| |||
Db   1373 CTACTGACACAGTCATGACTCCGCCCATTGCCTGTGTTAAGCCCAAGGTTAAGCAGA 1432

Qy   1460 AAGTGAGAATAAAAGACTGTGTAGTTCGTCCCCAGCCGGCATACCCTTCTTTATGTC 1519
          ||||| || ||| |  |||| ||  || ||  ||||| |||  ||| ||||||  |
Db   1433 GGGTGAGGTGAGAACAGTGCCTTCTCTTCTGTTCCAGGCTGGCAACTTCCCTGGT 1492

Qy   1520 AG------------TGAACGGTGACCAGAGTGATGTTTGCAATTAGTGGGGCCAAGAC 1564
          ||              |   || || || |||| || || ||  || |||  ||||
Db   1493 CGTGAGAACAGTGAGTGTCGACCAAATGGCCAACACATGCCGCACTTGTGAGGTCAAGGA 1552

Qy   1565 CTGGCTAACCGGCCAAGTGTTTTTGGGAGCAGCGCCCCACTTTCGGGACCAAGAC 1608
          |||  ||  ||  | |||| || ||||| || ||| | || |||  || || |||
Db   1553 CTGTCCAGGCAGCGGGTTTTGTGGGACAGAGCGCCCCACTTTCCTGTGTTCACTG 1612

Qy   1609 --------------AGTCCAAGAGAGAGATCCCCAGAAAACACTTCCTGTGTTCACTG 1652
                        || ||| || || |||||||||||||| |||||||| |||||||
Db   1553 CTGTCCAGGCAGCGGGTTTTGTGGGACAGAGCGCCCCACTTTCCTGTGTTCACTG 1612
```

```
RESULT  6
BC000310
LOCUS       BC000310               2010 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, TRAF interacting protein, mRNA, complete cds.
ACCESSION   BC000310
VERSION     BC000310.1  GI:12653088
KEYWORDS    .
SOURCE      MGC.
            human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2010)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site:   http://www.nisc.nih.gov/
            Contact:    nisc_mgc@nhgri.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L.-H. and Green,E.D.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 1 Row: e Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 5032194.
FEATURES             Location/Qualifiers
     source          1..2010
                     /organism="Homo sapiens"
                     /db_xref="LocusID:10293"
                     /clone="MGC:8424 IMAGE:2821007"
                     /db_xref="taxon:9606"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             97..1506
                     /codon_start=1
                     /product="TRAF interacting protein"
                     /protein_id="AAH00310.1"
                     /db_xref="GI:12653089"
                     /translation="MPIRALCTICSDFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSR
                     TCPQCRIQVGKRTIINKLFFDLAQEEENVLDAEFLKNELDNVRAQLSQKDREKRDSQV
                     IIDTLRDTLEERNATVVSLQQALGRAEMLCSTILKQMKYLEQQDETKQAQEEARRLR
                     SKMKTMEQIELLLQSORPEVEEMIRDMGVGQSAVEQLAVVCVSLKKEYENLKEARKAS
                     GEVADKLRKDLFSSRRKLQOTYYSEDLQAKLELKSAQKDLQSADKEIMSLKKKLTMLQE
                     TLNLPPVASETVDRLVLESPAPVEVNLKLRRPSFRDDIDLNATFDVDTPARPSSSOH
                     GYYEKLCLEKSHSPIQDVPKKICKGPRKESQLSLGQOSCAGEPDEELVGAFPIFVRNA
                     ILGQKQPKRPRSESSCSSKDVVRTGPDGLGGRTKFIQPTDVYMIRLPVKPKTKVQRV
                     RVKTVPSLFQAKLDTFLWS"
BASE COUNT      517 a    519 c    561 g    413 t
ORIGIN
```

```
Query Match
Best Local Similarity   54.3%;
Matches 1458;   Conservative    0;   Mismatches  428;   Indels    59;   Gaps    7;
                        75.0%;   Pred. No. 3e-296;   DB  9;   Length 2010;
                                        Score 1072.2;

Qy       81 TGGTTCCCCTGGCGTCTGCTTGACTGCGACCCATCATGCTCATTCCTCTCGTGCACTATCTG 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       66 TGGCTGCCTGGCCGCCCTTGAGTCCAGCCATCATGCTCATTCCTCTCGTGCACTATCTG 125

Qy      141 CTCCCGACTTCTTCGATCACTCCCGTGACGTGGCTGCCATCCACTGTTGGGCCACACTTTTCA 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      126 CTCCGACTTCTTCGATCACTCCCGCGACGTGGCCGCCATCCACTGCGGGCCACACTTTCCA 185

Qy      201 TCTGCAATGCCTAATCCAGTGGTTTGAGACAGCACCACCAAGTGGAGCTGCCACAGTCTAG 260
            |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      186 CTTGCAGTGCCTAATTCAGTGGTTTGAGACAGCACCAAGTGGAGCTGCCACAGTCTCG 245

Qy      261 AATCCAGGTTGGCAAAAAAGACTATATTAATAAACAAACATTTTCTTGACCTCGCCCCAGAAGA 320
            ||||||| ||||||||||||||||||| |||||| |||||||||||||||||||||||
Db      246 AATCCAGGTTGGCAAAAGAACCATTATCAATAAGCTCTTCTTTGATCTTGCCCAGGAGGA 305

Qy      321 GGAGAATGTCTTGGATGCAGAATTCTTAAAGAATGAACTGGACACGCTCAAAGCTCAGCT 380
            ||||||| |||||||||||||||||||||||||||||||||||||||| |||||||
Db      306 GGAGAATGTCTTGGATGCAGAATTCTTAAAGAATGAACTGGACAATGTCAGAGCGCCAGCT 365

Qy      381 TTCCCAGAAAGACAGGGAGAAGCGGGACAGCCAGGTCATTATCGACACTCTACGGGACAC 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      366 TTCCCAGAAAGACAGGGAGAAACGAGACAGCCAGGTCATCATCGACACTCTGCGGGATAC 425

Qy      441 CCTGGAAGAACGGCAATGCTACCGTGGAGTGTCCTCACGAACAGCCCTTAAACAGCGAGAT 500
            ||||||||| |||||| |||||||||||| || |||||||||||||||||||||||||||
Db      426 GCTGGAAGAACGCAATGCTACTGTGGTATTCTGCAGACAGCCCTTGGGCAGGGCCGAGAT 485

Qy      501 GCTGTGTTCCACCCTGAAAAAACAGATGAAGTTCCTGGACAGCCGGAGATGAGACCAA 560
            |||||||| |||| |||||||||||||||||||||||||||||||||||||||||||
Db      486 GCTGTGTCCACACTGAAAAAGCAGATGAAGTACTTAGAGCAGCAGGATGAGACCAA 545

Qy      561 ACAAGCTCGGGAGGAGGCCCGACGACTCAAGTGCAAGATGAAAAACCATGGAGCAAAATTGA 620
            |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      546 ACAAGCACAGAGGAGGCGCCCCGGCTCAGGAGCAAGATGAAGACCATGGAGCAGATTGA 605

Qy      621 GCTCCTACTCCAGTCCAGGCCGTTCTGAGGGTGGAGGAGATGATTCAGACACATGGGGTGGG 680
            ||| ||||||| ||||| |||||||||||||||||||||||||||||||||||||||||||
Db      606 GCTTCTACTCCAGAGCCAGCGCCCTGAGGTGGAGGAGATGATCCGAGACATGGGTGTGGG 665

Qy       81 ACAGTCAGCGCGTGGAACAGCTGGCGTGTTGTACTGTGTCTCTCAAGAAAAGAGTACGAGAA 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      666 ACAGTCAGCGAGTGGAACAGCTGGCTGTTGTACTGTGTCTCTCAAGAAAAGAGTACGAGAA 725
```

```
Qy     1613 GGCAGAGAAACAGGTGAGGATGAGGTGTGGCCACCCAGAGACTGCTCTTCCTGCCCTCACC 1672
            ||| | |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     1653 CGCCCTGCACC--ACACTGGGGAAGCCACACAGTTTACTGTTCCGATCAGCAGGGCC 1710

Qy     1673 TGCCCACTCCTACGACTGGGAGCTGACATGACCAACCACCCACTGATCCTGTGCAGCTCC 1732
            |||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     1711 TACTTCCAGTTGCCAGGTTTTGCTTATAGCTACAACACCAGGTGTGGGCTGCCTTTTGT 1770

Qy     1733 TGCTCCTGTTGCCAGGCTCCTCGTTTATAGCCATGATCAGATGTGTCAGACTCTTTCTGG 1792
            |||||||| ||||||||||||| |||||||||||||||||| |||||||||
Db     1771 TTTTATAGAACAGGGTGCACATTGACTCTAAGTGGATGGGAGTGGCTGGAGGATCCTATGCA 1830

Qy     1793 GCCTGGAGGACCACCAGGTCACTTGTTGACTGTGCTGTCTGTGGACCAG-----AGTGCTTGAGGC 1847
            |||||||||||| |||||| || ||||||||||||||||||||||||  |||||||||
Db     1831 GGCTGGAGGACCCTGCGCTGCCTTGAAACTCCTGCCTGCCCACCTTTATTGCTTATTATGG 1890

Qy     1848 ATCTCAGGCAGCCCTCAGCCCAAGCTTCTACCTGCCCTTTGACTTGCTTCTAGGCATAGCCT 1907
            |||||||||||||||||||||||||||||||||| ||||||
Db     1891 GGTGAGGTGGTGATAGGGAAAGGTTGGGGAAGTTTTTCTGTGTAAAATAAAAAAGGGATCTT 1950

Qy     1908 GGGCAAGCAGGGTGGGTAATGAGGAGGATACCATGGGATGTTATGGAGAGACGATGGGAAGAGATTT 1967
            | |||||| ||||||||||||||||||||||||||
Db     1951 TTTCTTCAAAAAAAAAAAAAAAA 1975

Qy     1968 TCATGTAAAATAAAAAATTAAAAAAAA 1992
            ||||||||||||||||||||
```

```
Qy   741 TCTGAAGGAAGCTCGGAAGGCCACAGGGGAACTGGCTGACAGGGGTTGAAGAAGGATTTGGT 800
         ||| || || |||||||| |||||||  ||||||    |||  ||||||| |||| |||
Db   726 TCTAAAAGAGGCAGGAAGGCACGGAAGGCCCTCAGGGGAGGTGGCTGACGAAGTTGAGGAAGGATTTGTT 785

Qy   801 GTCCTCTAGGAGCAAGAGTTGAAGACTCTCAACACTGAGCTGGATCAGGCCAAGTTAGAACT 860
         || | ||||||||  ||||||||||  |  ||||| |||||| || || ||||||||||||
Db   786 TTCCTCCAGAAGCAAGTTGCAGGACACAGTTGGATCAGTTGGATCGGATCCAAGTTAGAACT 845

Qy   861 GAGGTCAGGCGACCCAGAGGATCAGCCTTACAAAAGTGCTGACAGATCACGAGAGAAAAGAA 920
         || ||||| | ||| |||||| || |||| || |||  |||||| || ||||||  ||||
Db   846 GAAGTCAGCCGACGCCCAGAGAGGACTTACAGAGTCTGACAGATCATGAGAGCCTGAAAAGAA 905

Qy   921 GTCCTGATGATCCTCCAGGGAAACCTTGAGCCTGCTCC-CCGACCAATGACGACGGTCAGCC 979
         |  || ||| |||| ||||||||||| ||||| || | |||||| ||| |  |||||||| ||
Db   906 G-CTAACGATGCTCCAGGGAAACCTTGAACCTGCTGGCCAGTCATTGTGAGACTGTCGACC 964

Qy   980 GGCCTGGTGTTTTGGAGAGCACAGGGCCCCTGTGGAGATGAATGAACCGAGGCTTCACAGCCAC 1039
         ||||||| |||| |||||| || ||||| |||  ||||| || ||||| || |||||||| |
Db   965 GGCTGGTTTTAGAGGACGAGCCCAGCCCTGTTGA--GGTGAATCTGAAGCTCCGCGCGGCCAT 1021

Qy  1040 CCTTCGGTGATGAGATTGAAGACTTCAATACCACCTTTGATGTAAATACCGCGCTCCAACCCAGA 1099
         |||||||| || | |||  |||| | |||| ||  ||| |||  ||||| ||  || |||||||
Db  1022 CCTTCGGTGAAGAGATTGATCTCAATACCACCTTTGATGTAAATACCACCTTCCAACCCAGA 1081

Qy  1100 CCTCTGGCTCTCCACCAATTGCCTCCCCACGAAAGCTTGTGTGCCTGGACAGGGCACGCGCTCTCCCA 1159
         ||| || |||||| || || | || ||| |  ||| ||||| ||||||||||| ||||| ||||||||
Db  1082 CCTCCAGCTCCCAGCCACATTGGTTACTACGAAAAAACTTTGCCTTAGACAGTCACACTCCCCA 1141

Qy  1160 TGCAGAANTGCTCATCAAGAAGAGTGCACAAAAGTCTCCAAGCCGGAGTCCCAGCTCTCACTGG 1219
         |||| || |  |  | ||||||  | ||| |||  | |||||||||| |||||||||||||||| |
Db  1142 TTCAGGATGTCCCAAGAAGAAGATATGCCTAAAGGCCGGAGGAAGGAGTCCCAGCTGTCAGTTGG 1201

Qy  1220 GTGGCCAGGACGATGTGTAGGAGGAGGAGGGAGGGACACGGCGCTCTCCA 1279
         ||||| || ||||| |  |||||||||| |  ||| ||||| || |||| |
Db  1202 GTGCCCAGAGCTGTGCCAGAGATGGATGAGAAAGGACCCAGAGGAACTGGTTGGTCGTTGCCTTCCCTACTTTG 1261

Qy  1280 TCCGGAATGCTGTCTGCTGGGTCAGAAAACAGCCCAAGCA 1339
         ||| ||||||| |  |  ||||| | |  |||||||||||
Db  1262 TCCGGAATGCCATCTAGGCGCGAAAACAGCCCAAGCTCGACAGTCTGAGATGTCATCCTCTTGCA 1321

Qy  1340 GCACAGAGTGTGGGTAAGAATAGGCTTTGATGGGGCTTGGAGGACGAACAAAATTCATTCCAGC 1399
         |||| ||| ||||| |||| |  |||| |||| |||  ||  |||| || |  |||||||||||
Db  1322 GCAAAGAGTGGGTTAAGGACAGGCCTTCGATGGGCTGCCGGGCGACAAAATTCATTCCAGC 1381

Qy  1400 CTAGGGACACAACCATTATCGCACCAGTGCCTGCTGTTTGAAGTCCCAAAGGCCTAAAACAGA 1459
         |||| || || || ||| | || || ||||| |||||||||| |||||||||||||||||||
Db  1382 CTACTGACACAGTCATCTGCCCCAATGGCCTGCTGTTTGAAGCTCCAAGCAAGGTTTAAGCAGA 1441

Qy  1460 AAGTGAGAATAAAAGACTGTGTCAGTTCTGTGCCTCCCAGCTGATAGCTCTTTATGTC 1519
         |||| ||| |  ||  | | || |||  | || |||||| || || |||||  |||| |
Db  1442 GGGTGAGGGTGAAGACAGTGCCTCTCTTTCCAGGCCAAGCTGGACACCTGTGGT 1501

Qy  1520 AG------TGAACGGTGACCAAGTGAGTGTTGCAATTAGTTCGGGCCAAGAC 1564
         ||      || | ||||  ||||| ||| | |||| | || |  ||||||||
Db  1502 CGTGGAGAACAGTGACTCTGACCAATGGCCAGACACATGCCTCAACTTGTTGTAGGTCAAGGA 1561

Qy  1565 CTGGGTAACGGGAAGTGTTTTTGGAAGATGGGCTCTCTGTGGACC------ 1608
         ||||  |||  | ||||| | |||  | ||| ||  |   |||||
Db  1562 CTGTCCAGGCAGCGGGTTTTGTTGGACGACAGAGCCCCACTTCCGGGACCCTGAGGTGTAAG 1621

Qy  1609 ------------AGTCCAAGAGAGAATGCCCAGAGAAAACACACTTCCTGTGTTCACTG 1652
                     |||||||| |  ||||| |  ||||||| | | ||||| |  |||||||
Db  1622 GGCACAGACAAACAGGTGGAGGATGAGGTGGCACCCAGAGACTGCTCTTCTGCCCTCACCC 1681

Qy  1653 CGGCCCTGCCACC---ACACTGGGGAAGCCACCATGACCAGCAGTTTACTGTTCCGATCAGCAGGGCCC 1710
         ||||  | | ||    ||| | ||||||| |||| | ||| ||||  | | |||  || |||| |
Db  1682 TGCCCCACTCCTACGACTGGGGAGCGTGACAATGGCCCACTGATCCTGTGTCAGCAGGTCC 1741

Qy  1711 TACTTCCAGTTTGCAGGGTTTTGCTTATAGCTACAACCAGGTCTACAACGGTTTGTT 1770
         |||||||||| || |||||||||  | | | |  | |||| ||| | |||| ||||| ||
Db  1742 TGCTCCTGTTGCCAGGCTCCTGTTTATAGCCATGATCAGAGATGTGGTCAGACTCTTTCTGG 1801

Qy  1771 TTTTATAGAACAGGGGTCACATTGACTCTAAGTGGATGGGAGAGTGCTGCTGGAGGATCCTATGCA 1830
         || | || || ||||| | ||| | |  |||| |||||||| |||||| | |      ||||| ||
Db  1802 GCCTGGAGGACCACGGGTCACTTGTTGACTGTGTCTGGTGGGAGTGCTTGAGGC-----AGTGCTTGAGGC 1856

Qy  1831 GGCTGGAGGACCCTGGCGCTTGAACTCCTGCCTGCCTCCAGCTTGTTGAAATTATGG 1890
         ||||| ||||| |||||| |||||| || || ||||  || |||  |  ||||| ||
Db  1857 ATCTCAGGCAGCCTCAGCCCCAAGCTTCTTCACTCTGCCTTTGACTTGCTCTAGGCCT 1916

Qy  1891 GGTGAGGTGGTGATGAGGGAAAGGTTGGGGGAAGTTTGGGGTAAAATAAAAAAGGGATCTT 1950
         || || || |||| || | |||||||| |||||  | | |||  |||  ||||||||
Db  1917 GGGCCAAGCAGGGTGGGTGGGTAATGGGAGGATAGCATGGGGAGGTGAGGATTT 1976

Qy  1951 TTCTTCAAAAAAAAAAAAAAAAAA 1975
         | | |||||||||||||||||||
Db  1977 TCATGTAAAAATAAAAAATTAAAAAAAAA 2001
```

```
RESULT  7

HSU77845

LOCUS       HSU77845      2007 bp    mRNA    linear   PRI 25-APR-1997
DEFINITION  Human hTRIP (hTRIP) mRNA, complete cds.
ACCESSION   U77845
VERSION     U77845.1  GI:2039303
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2007)
  AUTHORS   Lee,S.Y., Lee,S.Y. and Choi,Y.
  TITLE     TRAF-interacting protein (TRIP): a novel component of the tumor
            necrosis factor receptor (TNFR)- and CD30-TRAF signaling complexes
            that inhibits TRAF2-mediated NF-kappaB activation
  JOURNAL   J. Exp. Med. 185 (7), 1275-1285 (1997)
  MEDLINE   97258620
REFERENCE   2  (bases 1 to 2007)
  AUTHORS   Lee,S.Y., Lee,S.Y. and Choi,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1996) Immunology, The Rockefeller University,
            1230 York Avenue Box 295, New York, NY 10021, USA
FEATURES             Location/Qualifiers
     source          1..2007
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..2007
                     /gene="hTRIP"
     CDS             103..1512
                     /gene="hTRIP"
                     /codon_start=1
                     /product="hTRIP"
                     /protein_id="AAB52993.1"
                     /db_xref="GI:2039304"
                     /translation="MPIRALCTICSDFFDHSRDVAAIHCGHTFHLQCLIQSFETAPSR
                     TCPQCRIQVGRRTIINKLFFDLAQEEENVLDREFLKNELDNVRAQLSQKDKERDSQV
                     IIDTLRDTLEERNATVVSLQQALGRAEMLCSTLKQOMKYLEQQODETKQAQEEAGRLR
                     SKMKTMEQIELLLLQSOLPEVEEMIRDMGVGQSAVEQLAVCVSLKKEYENLKEARKAS
                     GEVADKLRKDLFSSRSKLQTYSEDLQAKLELKSAQKDLQSADKEIMSLKKKLTMLQE
                     TLNLPPVASEIVDRLVLESPAPEVENLKLRRPSFRDDIDLNATFPVDTPARPSSSOH
                     GYEYKLCLEKSHSPIODVPKKICKGPRKESOLSLGGOSCAGEPDEELVGARPIEVRNA
                     ILGQKOPKPRSESSCSKDVVRTGFDGLGGRRKFIQPTDIVMIRPLPVKPKTKVKQRV
                     RVKTVPSLFQAKLDTFLWS"

BASE COUNT      517 a    518 c    558 g    414 t
ORIGIN
```

```
Query Match        53.8%;  Score 1062.8;  DB 9;   Length 2007;
Best Local Similarity  74.3%;  Pred. No. 1.5e-293;
Matches 1498;  Conservative 0;  Mismatches 457;  Indels 61;  Gaps 10;

Qy    9 GTGCGGTGGGAGCGAAATTTGAAGGAACCGGACGGTGGCCGGTTCCACCAAAACTGTCT 68
         |||||||||||||||||||||||| |||| ||||| |||||  ||   |||  |||| |
Db    1 GTGCGGTGGGAGCGAAATTTGAAGCAACGCGGAGGCGGGGGC--CCTCTACGGAAGCCGGACCT 58
```

```
Qy     69 GTCTCTG-GCAGCTGCGTTCCCTGGGCTGCTTGAGTCGGACGCCATCATGCCTATCCTCTC  127
          ||| ||    ||  ||  |||||| |  ||  |||||| ||  |||||||||||||| | ||
Db     59 GTAGCAGTTGTTCTTTGGCTGCTGGGGCGCCCTTGAGTCCAGCCATCATGCCTATCCCGCTC 118

Qy    128 TGTGCACTATCTCGCTCCGACTTCTTCGAGTCACTCCCGTGACGTGCCATCACTCCCGTG  187
          |||  || | |||  | |||| ||||| ||||| | | |||  || |||||||| ||| |
Db    119 TGTTCCACTATCTGCTCCTTCGATCACTCCGGTGACGTGGCCATCCACTCACTCGG  178

Qy    188 GCCACAACTTTTCATCTGCAAATGCCTAATCCCAGTGGTTTGAGACAGCACACCAAGTCGGACCT  247
          ||| || ||  |  ||||||  |  ||| ||||||||| |||||||||||| |||||||||
Db    179 GCCACACCTTCCACTGCAGTCAGTGCCTAATTCAGTCCTTTGAGACACCACCAAGTCGGACCT 238

Qy    248 GCCCACAGTTGTAGAATCAGGTTGGCAAATGATCATTATTAATAAAACATTTTCTTTGACC  307
          ||||||||||  |||| ||||  || |||| | || |||| |  || | ||||||||| |
Db    239 GCCCACAGTGCCGAATCCAGGTTGGCAAAAGAACCATTATCAATCAAGCTCTTCTTTGATC 298

Qy    308 TCGCCCCAGGAGGAGAAGAATGTCTTGGATGCAGAATTCAGGCAGAAGGACACGCG  367
          | ||||||||||||||| |||||||||||| |||||||||||||||||||||||
Db    299 TTGCCCAGGAGGAGAATGTCTTGGATCGAGAAATGCTGGACACAAATG  358

Qy    368 TCAAAGCTCAGCTTCCCAGAGAAAAAGACACAGGGAGGAGAAACGGAGGCCAGGTCATTATCGACA  427
          | |||||||||||||||||| || | ||||| |||||| |||||||||||||||||| |||||
Db    359 TCAGAGCCCAGGCTTTCCCAGAGAAGAAAGACAAGGGAGGAGAAACGGAGGTCATCATCGACA  418

Qy    428 CTCTACGGGACCACCCTGGAGAAACGGCAAATGCTTACCGGTGAGTCCCTACAAAACCCTTAA  487
          |||| ||||||||||| |||||||||||| ||||| |||||||||| ||  | || |||||
Db    419 CTCTGCGGGATCACCCTGGAGAAACCAATGCTTACTGTGTGGTATCTGCAGCACGGCCTTGG 478

Qy    488 ACAAGGCACAGAGATGCTTGTGTTCACCCTGAAAAAAACAGAGTGAAGTTCCTGGACGCAGCGC  547
          || ||| || ||||||||||||||||| || ||||||| ||| ||  ||  |||||| |||
Db    479 GCAAGGCCCAGAGATGCTGTCTCACACTGAAAAAGCAGATGATGAACTACTTTAGAGCAGCAGC 538

Qy    548 AGGATGAGACCAAAACAAAGCTCGGGAGGGAGGAGGCCCACGACTCAAGTGCAAGATGAAAACCA  607
          || |||||||||| | ||||||||||| ||||||||||||||||||||| | ||| ||||||
Db    539 AGGATGAGACCAAAACAACACACAGAGGAGGAGGCGGCGGGACTCCAGAGCAAGAATGAAGACCA 598

Qy    608 TGGAGCAAAATTGAGCTCCTACTCCAGAGCTGACTCCGTTTCTGAGGTGGATGATTCAGG  667
          |||||| |||||||||||  |  ||||| |||  || ||| ||||||| |||||||||
Db    599 TGGAGCAGAAATGAGCTTACTACGACACTGAGCTGGACGAGGTGGGAGGATGATGATTCAGG 658

Qy    668 ACATGGGTGTGGGACAGTCAGCGCTGGACCAGCTGCTGTGTCCCTCAAGAGCAGCAGGCGC  727
          |||||||||||| |||||||| ||||||||||||| || ||||| |||||||| || ||
Db    659 ACATGGGTGTTCACCGGTCAGCGCTGGACCAGCTCCCTGTGTGGTACTCTGCAGCACGGC 718

Qy    728 AAGAGTAATGAGATCTAAAGAGCCACACGGAAGCGGAGGGCCCAGGTTGA  787
          ||||| || || |||||||| |||||| || || ||||||||||||| ||
Db    719 AAGAGTACCACGAATCTAAAGAGGCCTCAGGGAAACCTTGAACCTGCCAGT  778

Qy    788 AGAAGGATTTGGTGTCCCTAGGAGCACACTGAGCTGGACTCAGG  847
          || ||||||| ||| || |||||||| |||||||||||||||
Db    779 GGAAGGATTTGTTTTCCTCCGGAAGCAAGTTGCGAGACAGCTCTACTTTGAAATTGGATCAGG 838

Qy    848 CCAAGTTAGAACTGAGGTCAGCCAGGACTTACAAAGTGCTGAACGATCACAGA  907
          || ||| |||||||| || |||||| |||| ||| |  ||||| |  |||||
Db    839 CCAAGTTAGAACTGAAGTCAGCCCCAGAAGGACTTACAGAGTGTGACAAGGAAATCATCA 898

Qy    908 GCCTAAGGTCAGCCGCCTGGTTCTGGAGCCCAGCCCCTGTGTGAGATGATGAACCCGAGG  966
          ||||  ||| |||| ||||| |||||| ||||| | || || || || | |||||||||
Db    899 GCCTGAAAAAGAAG-CTAACGAATGCTGCAGGAAACCTTGAACCTGCCACCAGTTGGCCAGT 957

Qy    967 GAGACGTTGCGACCCGCCGGTTCTGGTTTTTTGGAGAGCCCGACGCGAGG  1026
          || ||||| | || ||||  || ||| || ||||||| | | |||  |||
Db    958 GAGACTGTTGAGCCGGCCGGTTGGTGTTTTAGAGAGCCCCAGCCGCTGGAAGTCGGAAG 1014

Qy   1027 CTTCACCAGCCACCCTTGGGTGCAGAGTTGATCTCAAGACCTTTTGATGTGAAATACC  1086
          |||||||| | || |||| |  ||||||| |||||||| |||| |||||  |||||||
Db   1015 GGAAGGATTGTTTTCCTCCGTGATTGATCTCAACACCTTTGATGTGTAAATAACC 1074

Qy   1087 CCTCCAACCCAGACCTCTGGGCTCCCCAGGAAGAGCTGTGCCTGGGAGAGG  1146
          |||||||||||||||||||||||| |||||||||| ||||||||| ||||||
Db   1075 CCCCAGCCCGGCCTCCAGCTCCCAGGAACCTTTGCCTAGAGAAG 1134

Qy   1147 GCAGCCTCTCCCATGCAGAAATGTCCTCAAGAAGTGCACAAAGTCTCAAGCCCGAGTCC  1206
```

```
Db   1135 TCACACTCCCAAATTCAGGATGTCCCACAGGGAGGATATGCAAAGGCCCCAGGAAGAGTCC  1194

Qy   1207 CAGCTCTCACTCGGGTGCCCAGCAGCTGTGGCAGGATGTCCCAGAAGAGGAACCACAGCA  1266
          |||| | |||| || |||||||| ||||||||||| || || |||| |||||||| ||
Db   1195 CAGCTCCACTCTGGTGCCCAGCAGCTGTTGCAGGACGTAGGAGAGGAGAACAGGAGTCA  1254

Qy   1267 TTCCCTCTTCATCCGGAATGCTGTCCTGGGTCTCCAGGATCAGGAACCCCCAACAGGACCACA  1326
          |||| ||||| | || ||||| || |||||| |||| ||||| || || |||||| |||||
Db   1255 TTCCCTATTTTGTCCGGAATCCATCCTAGGCACCAGAAAACAGCCCCAGCCAAAAAGCCCTCA 1314

Qy   1327 GAATCCCGAAGCAGCACAGATGTGGTAAGAAATAGGCTTTGATGGGCTTGGGAGGACGAAACA  1386
          || | | |||| || ||||||||| |||| ||||||||||||||||| ||| ||||| ||
Db   1315 GAGTCCTCTTGCAGCAAAGAATGTGGTTAAGGAACAGGCTTCGAGGTGGGCCGGACA 1374

Qy   1387 AAATTCATCCAGCCTAGGGACACACCATTATCCGACAGCACCACCGCGTCCTCTTAAGTCGTGG  1446
          |||| ||||||||||| ||| |||| |||||| |||||| |||||| ||||| |||||||
Db   1375 AAATTCATCCAGCCTACTGACACAGTCATGATCCGCGCCCTGTTCAACCCCAAGACC 1434

Qy   1447 AAGAGTAAACAGAAAGTGGAGAATAAAAGACTGAGTAGTTCTGCCTCCCCAGCGCCCAAGCTGGAT  1506
          || || |||| |||| |||||||||| ||||||||||| ||  |||||||||||||||||
Db   1435 AAGGTTTAAAGCAGAGAGGGGTGAGGGGTGAAGACACGTGCCTTCTCTCTTCCAGGCTGGAC 1494

Qy   1507 ACCTTCTTATGTCAG----------------TGAACGGTGACGAGTGATGTTTACTGTTCCGA  1551
          ||| | | ||||||                ||||  || ||||||| ||||| ||||||
Db   1495 ACCTTCCTGTCGGTCGGTCGGTGGAACAGTCTGAGTCTGACCAATGGCCAGAACATGCCTGATCCTG 1554

Qy   1552 AGTGGCCCACAGAACCTGGCTAACCCGGG--------------AAGTG  1581
          |  || |||| | || |||| || ||||||              ||||
Db   1555 TGTAGGTCAAGGACTGTTCCAGGCAGTGTTTGGACAGGACCCTCTACTGTTTCCTTTCGGACCAGC 1614

Qy   1582 TTTTTGGAAGATGGCTCCTCTTGGACCAGTCACATTGACTCTAAGTGGATGGAGTTGCTGGAG  1641
          ||  |||||  ||| || ||  ||||||||| | | ||| || ||||| ||||||||| ||
Db   1615 TGAGGTGTAAGGGCACAACAAACAGTGAGGGTGAGTGTGACACCCAGACTGCTCTTCC 1674

Qy   1642 TGTGTTCACTGCGGCCCTGCACC--ACACTGGGAAGGCCACATGGACCACAGTTTACTGTTCCGA  1699
          ||||| ||| | |||||| |||  |  || |||||| |||||| |||||||||||| ||||
Db   1675 TGCCCTCACCCTGTCCCCACCTTGTCTACGGAGCTGAGTCTGAAGCAGACACGTTGGGAAGTTGAACCTTG 1734

Qy   1700 TCAGCAGGGGCCTACTTGCCAGTTTGCAGTCGGCTTTTAGTTATAGCTGACTGGAGGCTG  1759
          | ||||| | ||||| |||| || |||||||  | || |||| |||  ||||||  ||||
Db   1735 TCAGCAGGTCCTGC--TCTGTTTGCCAGCTCTTGTTGTTTTTTAGGCCATGAACAGATCAGATGGTCAG 1793

Qy   1760 ACTCCCTTTGTTTTTATAGAACAGGGTGTCACATTGACTCTAAGTGGATGGAGTTGCTGGAG  1819
          || | ||||| ||| ||||  ||||| || ||| |||| || ||||  |||||||| ||||
Db   1794 ACTCTTTTCTGGGGCGTCTGGGGAGGACCCGGTCCACTGGTTGTTGACTGTGTCTGCTGTGTGCCA 1848

Qy   1820 GATCCTATGCAGGGCTGGGAGCACCTGCGCTTGAACTCCTGCCTCCAGCTTATTGTGCT  1879
          | ||| |||||||| ||||| ||||||| |||||| |||||||| ||||| ||||||
Db   1849 GTGCTTGAGGCATCTCAGGGCGACCTCAGCCGCTTACCCAAGCTTCTACCTGTTTGGCTCT 1908

Qy   1880 TGAAATTATGGGGTGAAGGTGGTGGATGAGGACGTTGGGGAGAAAGGTTTCTGTGTAAATAA  1939
          ||||| | || | |||||| ||| ||||||||| | || || ||||| || |||| || ||
Db   1909 AGCATAGCCTCGGGCCAAGCAGGTGGGAATGGGAAGCAGGATAGACATGGGATGTATGGAGAGG 1968

Qy   1940 AAAGGGGATCTTTTCTTCAAAAAAAAAAAAAAAAAAAAAAAAA  1975
          |||||| | |||| |||| |||| |||||||||||||||||
Db   1969 ATGGAAGATTTTCCGGAAAAAAAAAAAAAAAAAAAAAAA  2004
```

AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (27-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 16 Row: a Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 2039305.
FEATURES   Location/Qualifiers
     source   1..3694
                     /organism="Mus musculus"
                     /db_xref="LocusID:22036"
                     /db_xref="taxon:10090"
                     /clone="MGC:11463 IMAGE:3156502"
                     /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   277..948
                     /codon_start=1
                     /product="TRAF-interacting protein"
                     /protein_id="AAH06929.1"
                     /db_xref="GI:13905262"
                     /translation="MPIRALCTICSDFDHSRDVAAIHCGHTFHLQCLIQWFETAPSR
                     TCPQCRIQVGKKTIINKLFPDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQA
                     IIDFLRDTLEERNATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLK
                     CKMKTMEQIELLLGSQREVEEMIRDMGVGQSAVEQLAYCVSLKKCVSSGVPLPSSF
                     VMQIL"
BASE COUNT   858 a   951 c   976 g   909 t
ORIGIN

Query Match   48.2%; Score 952.6; DB 10; Length 3694;
Best Local Similarity   93.1%; Pred. No. 6.6e-262;
Matches 1057; Conservative   0; Mismatches   9; Indels   69; Gaps   3;

Qy   902 TCACGAGCCTAAGAGAAAGAGAGTCTGATGAATGATCTCCAGGGAACCTTGAGCCTCCTCC-GCG 960
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   2561 TCCAGAGCCTAAGAGAAGAAG-CTGATGAATCCTCCAGGGAACCTTGAGCCTTGAGCCTCCGGGCG 2619

Qy   961 ACCAATGAGACGGTCAGCCCGCCGCCTGGTTTTTGAGAGCCCAGCCCCTGTGGAGATGATGAAC 1020
            || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2620 ACCAATAGAGACGGTCAGCCCGCCCTGGGTTTTTGAGAGCCCAGCCCCTGTGGAGATGATGAAC 2679

Qy   1021 CCGAGGCCTTCACCACCACCCCTTCGGTGATGAGGATTGAATCTCAATACCACCTTTGATGTA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2680 CCGAGGCCTTCACCACCACCCCTTCGGTGATGAGGATTGAATCTCAATACCACCTTTGATGTA 2739

Qy   1081 AATACCCCTCCAACCCAGACACCTCTCGGCTCTCCCAAGAGAGCTGTGTCCCTG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2740 AATACCCCTCCAACCCAGACACCTCTCGGCTCTCCCAAGAGAGCTGTGTCCCTG 2799

Qy   1141 GAGAGGGCAC---------------------------------- 1150
            ||||||||||
Db   2800 GAGAGGGCACACGCCTGCCAGTTGAGGCTGTGAATGCTCATCCTCTGTGCCATTGGC 2859

Qy   1151 -----------------GCTCTCCCATGCAGAAGTGTCCTCAAGAGAGGTGCACAAAGTCTC 1193
                             |||||||||||||||||||||||||||||||||||||||||||||||
Db   2860 TGGGCTCCCTAGGAACTTGTCTCCCATGCAGAAGTGTCCTCAAGAGAGGTGCACAAAGTCTC 2919

Qy   1194 CAAGCCGGAGTCCCAGCTCTCACTGGGTGGCCAGCCGGATGTGTAGGAGGAGTGAGGA 1253
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2920 CAAGCCGGAGTCCCAGCTCTCACTGGGTGGCCAGCCGGATGTGTAGGAGGAGTGAGGA 2979

Qy   1254 ACTGGCTGGTGCCTTCCCTCTCTTCATCCGAATGCTGGTGTCCTGGGGTGTAGGGAGAAACAGCCAA 1313
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2980 ACTGGCTGGTGCCTTCCCTCTCTTCATCCGAATGCTGGTGTCCTGGGGTGTAGGGAGAAACAGCCAA 3039

Qy   1314 CAGGACCACAGCAGCAGAATCCGAGCAGCACACAGAGATGTGGTAAGAAGTAGAGGGCT 1373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3040 CAGGACCACAGCAGCAGAATCCGAGCAGCACACAGAGATGTGGTAAGAAGTAGAGGGCT 3099

Qy   1374 TGGAGGACGAACAAAAATTCATCCCGTTGCAGCACAACAATTATTCCGAACCAGTCCTGT 1433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3100 TGGAGGACGAACAAAAATTCATCCCGTTGCAGCACAACAATTATTCCGAACCAGTCCTGT 3159

Qy   1434 TAAGTCCAAGGCCAAGAGTAAACAGAAAGTGAGAATAAAGACTGTCAGTTGTTGCCTCCCA 1493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3160 TAAGTCCAAGGCCAAGAGTAAACAGAAAGTGAGAATAAAGACTGTCAGTTGTTGCCTCCCA 3219

Qy   1494 GCCCAAGCCTGGATTACCTTCTTTAATGTCAGTGAACGGGTGGACCAGAGTGATGTTTGCAATTAG 1553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3220 GCCCAAGCCTGGATACCTTCTTTAATGTCAGTGAACGGGTGACCAGAGTCATGTTTTGCAATTAG 3279

Qy   1554 TGGGCCAAGACCTGGCTAACCGGAGAAGTGTTTTTGGAAGATGGCCTCTTCTTTGGACCAGTCC 1613
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3280 TGGGCCAAGACCTGGCTAACCGGAAGTGTTTTTGGAGAGATGGCTCTCTTCTTTGGACCAGTCC 3339

Qy   1614 AAGAGAGGATGCCCAGAAAACACACTTCCTGTGTTCAGTGCCGCCCTCCACCACACTGGGAA 1673
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3340 AAGAGAGGATGCCCAGAAAACACACTTCCTGTGTTCACTGCCGCCCTCCACCACACTGGGAA 3399

Qy   1674 GCCACATGACCAGTTTACTGTTCCGATCAGCAGGGCCTACTTCCAGTTGCAGTTTTGC 1733
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3400 GCCACATGACCAGTTTACTGTTCCGATCAGCAGGGCCTACTTCCAGTTGCAGTTTTGC 3459

Qy   1734 TTATAGCTTACAACCAGGTGTGGCCAGTTCCTGGACTCTCTTTTTTATAGAACACAGGGTCACATTG 1793
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3460 TTATAGCTACAACCAGGTGTGGCCTGGACTCCCTTTTGTGTTTTTATAGAACACAGGGTCACATTG 3519

Qy   1794 ACTCTAAGTGGATGGGAGTGCTGGAGGATCCTATGCAGGCTGGAGGAGGACCCTGCGGCTTGAA 1853
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3520 ACTCTAAGTGGATGGGAGTGCTGGAGGATCCTATGCAGGCTGGAGGAGGACCCTGCGGCTTGAA 3579

Qy   1854 CTCCTGCCTGCCTCCAGCTTATTGCTTGAAATTATGGGGTGAGGGTGGTGATAGGGAAAGG 1913
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3580 CTCCTGCCTGCCTCCAGCTTATTGCTTGAAATTATGGGGTGAGGGTGGTGATAGGGAAAGG 3639

Qy   1914 TTGGGGAAGTTTTCTTGTGTGTAAAATAAAAAGGGATACTTTCTTCAAAAAAAAAA 1968
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   3640 TTGGGGAAGTTTTTCTGTGTGTAAAATAAAAAGGCATATCTTTCTTCAAAAAAAAAAAAAAAAAA 3694

RESULT   9
LOCUS   AX350344
AX350344   7542 bp   DNA   linear   PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0192492.
ACCESSION   AX350344
VERSION   AX350344.1   GI:18616006
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Yan,C.C., Wei,M.H., Ketchum,K.C., Merkulov,G.C. and Beasley,E.M.
  TITLE   Isolated human kinase proteins, nucleic acid molecules encoding

JOURNAL human kinase proteins, and uses thereof
  Patent: WO 0192492-A 3 06-DEC-2001;
  Applera Corporation Robert A. Millman Assistant Secretary (US)
FEATURES Location/Qualifiers
 source 1. .7542
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
BASE COUNT 1612 a 1977 c 2156 g 1797 t
ORIGIN

Query Match 5.6%; Score 111.4; DB 6; Length 7542;
Best Local Similarity 76.8%; Pred. No. 2.1e-20;
Matches 162; Conservative 0; Mismatches 46; Indels 3; Gaps 2;

```
Qy      2 GCACGAGGTGCGGTGGAGCGAAAATTTGAAGGAAACCGAG--CGGTGGCGGTTCCACCAA   59
          |||| |||||| ||||| ||||||||||||||| ||||||   ||||| ||||||||||||
Db   7295 GGAAGTCGTGCGTGCGGGGAGGCCAAAATTTGAAGCAAAGCGGAGGCGGGGCCCGTTCACCAA 7354

Qy     60 ACTGTGTCTGTCTCTG-GCACGTGGTTCCCGTGGCCGCTTGAGTCGGAGCCATCATGCCTA  118
          | ||| |||||||||||  ||| |||||||| |||| |||||||||||||||||||| |||
Db   7355 GCCGGACCTGTAGCAGTTTCTTTGCTGCTCGGCCCCCTTGAGTCGGAGCCATCATGCCCTA 7414

Qy    119 TCCTCTCTCTGTGCACTATCTGCCTCCGACTTCTTCGATCACTCCCGTGACGTGGCTGCCA  178
          |||| ||||||||  |||||||||| |||||||||||| ||||||||||||||||||||||
Db   7415 TCCGTGGCTCTGTGCACTATCTGCCTCCGACTTCTTCGATCACTCCCGCGACGTGGCCGCCA 7474

Qy    179 TCCACTGTGTGGCCACACTTTTCATCTGCAATG  209
          |||||||||||||||||||||||||||||||||
Db   7475 TCCACTGCGGGCCGCCACACCTTCCACTTGCAGTG 7505
```

RESULT 10
LOCUS AC099330 163424 bp DNA linear HTG 09-NOV-2001
DEFINITION Homo sapiens chromosome 3 clone RP11-78010, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AC099330 AC068701
VERSION AC099330.1 GI:16874864
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163424)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163424)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (09-NOV-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Nov 9, 2001 this sequence version replaced gi:8247914.
COMMENT -------------- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: http://www.genome.washington.edu
 Contact: uwgchtgs@u.washington.edu
 Drafting Center: BCM
 -------------- Project Information
 Center project name: chr-3
 Center clone name: RP11-78010 (bc0618)
 -------------- Summary Statistics
 Sequencing vector: unknown; 42% of reads
 Sequencing vector: plasmid; 58% of reads
 Chemistry: Dye-terminator ET; 92% of reads
 Chemistry: Dye-terminator Big Dye; 8% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 158969 bases at least Q40
 Consensus quality: 160979 bases at least Q30
 Consensus quality: 162023 bases at least Q20
 Insert size: 202900; 21.9% error; agarose-fp

 Insert size: 162724; sum-of-contigs
 Quality coverage: 9.0x in Q20 bases; agarose-fp
 Quality coverage: 11.2x in Q20 bases; sum-of-contigs
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 contig of 4608 bp in length
 * 4608: gap of unknown length
 * 4609 contig of 5581 bp in length
 * 10289: gap of unknown length
 * 10290 contig of 6716 bp in length
 * 17105: gap of unknown length
 * 17106 contig of 27895 bp in length
 * 45100: gap of unknown length
 * 45101 contig of 19974 bp in length
 * 65174: gap of unknown length
 * 65175 contig of 25445 bp in length
 * 90719: gap of unknown length
 * 90720 contig of 28558 bp in length
 * 119377: gap of unknown length
 * 119378 contig of 43947 bp in length
 * 119477: gap of unknown length
 * 119478 contig of 43947 bp in length.
FEATURES Location/Qualifiers
 source 1. .163424
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="RP11-78010"
  /clone_lib="RPCI human BAC library 11"
  /chromosome="3"
 misc_feature 1. .4608
  /note="assembly_name:Contig74"
 misc_feature 4709. .10289
  /note="assembly_name:Contig75"
 misc_feature 10390. .17105
  /note="assembly_name:Contig76"
 misc_feature 17206. .45100
  /note="assembly_name:Contig77"
 misc_feature 45201. .65174
  /note="assembly_name:Contig78"
 misc_feature 65275. .90719
  /note="assembly_name:Contig79"
 misc_feature 90820. .119377
  /note="assembly_name:Contig80"
 misc_feature 119478. .163424
  /note="assembly_name:Contig81"
BASE COUNT 40950 a 40937 c 39779 g 40912 t 846 others
ORIGIN

Query Match 5.6%; Score 111.4; DB 2; Length 163424;
Best Local Similarity 76.8%; Pred. No. 2.4e-20;
Matches 162; Conservative 0; Mismatches 46; Indels 3; Gaps 2;

```
Qy       2 GCACGAGGTGCGGTGGAGCGAAAATTTGAAGGAAACCGAG--CGGTGGCGGTTCCACCAA     59
           |||| |||||| ||||| ||||||||||||||| ||||||   ||||| ||||||||||||
Db  149268 GGAAGTCGTGCGTGCGGGGAGGCCAAAATTTGAAGCAAAGCGGAGGCGGGGCCCGTTCACCAA 149327

Qy      60 ACTGTGTCTGTCTCTG-GCACGTGGTTCCCGTGGCCGCTTGAGTCGGAGCCATCATGCCTA    118
           | ||| |||||||||||  ||| |||||||| |||| |||||||||||||||||||| |||
Db  149328 GCCGGACCTGTAGCAGTTTCTTTGCTGCTCGGCCCCCTTGAGTCGGAGCCATCATGCCCTA 149387

Qy     119 TCCTCTCTCTGTGCACTATCTGCCTCCGACTTCTTCGATCACTCCCGTGACGTGGCTGCCA    178
           |||| ||||||||  |||||||||| |||||||||||| ||||||||||||||||||||||
Db  149388 TCCGTGGCTCTGTGCACTATCTGCCTCCGACTTCTTCGATCACTCCCGCGACGTGGCCGCCA 149447

Qy     179 TCCACTGTGTGGCCACACTTTTCATCTGCAATG    209
           |||||||||||||||||||||||||||||||||
Db  149448 TCCACTGCGGGCCGCCACACCTTCCACTTGCAGTG 149478
```

```
RESULT   11
HS465N24
LOCUS           HS465N24            129747 bp    DNA    linear   PRI 23-NOV-1999
DEFINITION      Human DNA sequence from clone 465N24 on chromosome 1p35.1-36.13.
                Contains two novel genes, ESTs, GSSs and CpG islands, complete
                sequence.
ACCESSION       AL031432
VERSION         AL031432.1  GI:4375969
KEYWORDS        HTG; CpG island.
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 129747)
  AUTHORS       Wilson,S.
  TITLE         Direct Submission
  JOURNAL       Submitted (11-FEB-1999) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
                requests: clonerequest@sanger.ac.uk
COMMENT         On Mar 7, 1999 this sequence version replaced gi:4176479.
                During sequence assembly data is compared from overlapping clones.
                Where differences are found these are annotated as variations
                together with a note of the overlapping clone name. Note that the
                variation annotation may not be found in the sequence submission
                corresponding to the overlapping clone, as we submit sequences with
                only a small overlap as described above.
                This sequence is the entire insert of clone 465N24. This sequence
                has been finished according to sequence map criteria as follows. An
                attempt is made to resolve all sequencing problems, such as
                compressions and repeats, but not necessarily within known
                annotated human repeat sequence elements (e.g. Alu). Where the
                sequence is ambiguous, there is an annotation using the 'unsure'
                feature key.
                This sequence was generated from part of bacterial clone contigs of
                human chromosome 1, constructed by the Sanger Centre Chromosome 1
                Mapping Group. Further information can be found at
                http://www.sanger.ac.uk/HGP/Chr1
                465N24 is from the library RPCI3 constructed at the Roswell Park
                Cancer Institute by the group of Pieter de Jong. For further
                details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..129747
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p35.1-36.13"
                     /clone="RP3-465N24"
                     /clone_lib="RPCI-3"
     repeat_region   540..753
                     /note="L2 repeat: matches 2255..2461 of consensus"
     repeat_region   975..1303
                     /note="AluJo repeat: matches 1..311 of consensus"
     repeat_region   1320..1547
                     /note="MIR repeat: matches 29..262 of consensus"
     repeat_region   2061..2268
                     /note="MIR repeat: matches 100..262 of consensus"
     repeat_region   2356..2518
                     /note="MER5A repeat: matches 5..164 of consensus"
     repeat_region   2519..2827
                     /note="AluYb8 repeat: matches 1..318 of consensus"
     repeat_region   3245..3360
                     /note="MIR repeat: matches 20..139 of consensus"
     repeat_region   3427..3485
                     /note="L2 repeat: matches 2688..2746 of consensus"
     repeat_region   3444..3676
                     /note="MIR repeat: matches 8..249 of consensus"
     repeat_region   3690..3856
                     /note="MIR repeat: matches 8..249 of consensus"
     repeat_region   4187..4494
                     /note="LTR41 repeat: matches 20..182 of consensus"
     repeat_region   4846..5143
                     /note="AluSg repeat: matches 1..308 of consensus"
     repeat_region   5287..5340
                     /note="AluSx repeat: matches 1..296 of consensus"
     repeat_region   5361..5518
                     /note="L1MD1 repeat: matches 5345..5398 of consensus"
     repeat_region   5508..6164
                     /note="L1MA4 repeat: matches 6152..6300 of consensus"
     repeat_region   6176..6245
                     /note="L1MD1 repeat: matches 5423..6091 of consensus"
     repeat_region   6247..6528
                     /note="35 copies 2 mer ag 67% conserved"
     repeat_region   6529..6674
                     /note="AluSg repeat: matches 15..291 of consensus"
     repeat_region   6741..6874
                     /note="73 copies 2 mer ta 73% conserved"
     repeat_region   7225..7278
                     /note="L1MD1 repeat: matches 6090..6223 of consensus"
     repeat_region   8137..8448
                     /note="MIR repeat: matches 36..85 of consensus"
     repeat_region   9092..9137
                     /note="AluSx repeat: matches 1..312 of consensus"
     repeat_region   9229..9326
                     /note="23 copies 2 mer at 100% conserved"
     repeat_region   9669..9732
                     /note="L2 repeat: matches 2574..2670 of consensus"
     repeat_region   10162..10365
                     /note="MIR repeat: matches 87..148 of consensus"
     repeat_region   10379..10505
                     /note="MIR repeat: matches 8..222 of consensus"
     repeat_region   10654..10814
                     /note="L2 repeat: matches 2586..2710 of consensus"
     repeat_region   11454..11752
                     /note="MRR86 repeat: matches 9..176 of consensus"
     repeat_region   11790..12171
                     /note="AluYb repeat: matches 1..302 of consensus"
     repeat_region   12172..12362
                     /note="L1M4 repeat: matches 1258..1640 of consensus"
     repeat_region   12363..12775
                     /note="MER53 repeat: matches 1..189 of consensus"
     repeat_region   14467..14600
                     /note="L1M4 repeat: matches 2742..2867 of consensus"
     repeat_region   14601..14885
                     /note="FLAM_C repeat: matches 1..133 of consensus"
     repeat_region   14893..15096
                     /note="L1M4 repeat: matches 2867..3278 of consensus"
     repeat_region   15159..15484
                     /note="MLT1D repeat: matches 286..505 of consensus"
     repeat_region   15487..16041
                     /note="FLAM_C repeat: matches 6..323 of consensus"
     repeat_region   16069..16773
                     /note="L1MC4 repeat: matches 7451..7973 of consensus"
     repeat_region   16774..17068
                     /note="L1ME3A repeat: matches 3211..3943 of consensus"
     repeat_region   17071..17384
                     /note="AluSq repeat: matches 3..297 of consensus"
     repeat_region   17385..17451
                     /note="AluSg repeat: matches 1..310 of consensus"
     repeat_region   17567..17775
                     /note="L1ME3A repeat: matches 3943..3955 of consensus"
     repeat_region   17912..18219
                     /note="L1ME3A repeat: matches 4012..4208 of consensus"
     repeat_region   18401..18922
                     /note="AluSq repeat: matches 1..313 of consensus"
     repeat_region   18923..19226
                     /note="L1ME3A repeat: matches 4315..4848 of consensus"
     repeat_region   14036..14330
                     /note="L1M4 repeat: matches 2224..2742 of consensus"
     repeat_region   14331..14466
                     /note="AluY repeat: matches 1..301 of consensus"
     repeat_region   11454..11752
                     /note="MIR repeat: matches 8..222 of consensus"
     repeat_region   11790..12171
                     /note="AluSx repeat: matches 1..304 of consensus"
     repeat_region   13476..14035
                     /note="L2 repeat: matches 1..291 of consensus"
     repeat_region   12833..13130
                     /note="AluJo repeat: matches 1..291 of consensus"
     repeat_region   12363..12775
                     /note="L1M4 repeat: matches 1640..2049 of consensus"
```

```
repeat_region    19227. .19530
                 /note="L1ME3A repeat: matches 4848. .4586 of consensus"
repeat_region    19531. .19897
                 /note="L1PB1 repeat: matches 5781. .6155 of consensus"
repeat_region    19898. .20698
                 /note="MIR repeat: matches 115. .213 of consensus"
repeat_region    21061. .21150
                 /note="L1ME3A repeat: matches 4585. .5403 of consensus"
repeat_region    21218. .21390
                 /note="L2 repeat: matches 1626. .1814 of consensus"
repeat_region    21511. .21767
                 /note="L1MC4 repeat: matches 6664. .6940 of consensus"
repeat_region    21771. .21969
                 /note="L1MC2 repeat: matches 3850. .4054 of consensus"
repeat_region    21970. .22008
                 /note="L1MC2 repeat: matches 4585. .5261 of consensus"
repeat_region    22009. .22354
                 /note="U2 repeat: matches 1. .39 of consensus"
repeat_region    22335. .22663
                 /note="AluSx repeat: matches 4054. .4410 of consensus"
repeat_region    22664. .23540
                 /note="L1MC2 repeat: matches 1. .309 of consensus"
repeat_region    23552. .23730
                 /note="L1MC2 repeat: matches 4410. .5261 of consensus"
repeat_region    23731. .24657
                 /note="FAM repeat: matches -135. .43 of consensus"
repeat_region    24659. .24758
                 /note="L1MC2 repeat: matches 5278. .6319 of consensus"
repeat_region    24759. .25029
                 /note="AluJ/FLAM repeat: matches 2. .77 of consensus"
repeat_region    25872. .26333
                 /note="L1MC4 repeat: matches 7461. .7777 of consensus"
repeat_region    26334. .26622
                 /note="L1MB5 repeat: matches 4928. .5368 of consensus"
repeat_region    26623. .26774
                 /note="L1MA8 repeat: matches 5995. .6287 of consensus"
repeat_region    26675. .27077
                 /note="L1MB5 repeat: matches 5368. .5512 of consensus"
repeat_region    27078. .27656
                 /note="AluSx repeat: matches 1. .298 of consensus"
repeat_region    27713. .27944
                 /note="L1MB5 repeat: matches 5512. .6176 of consensus"
repeat_region    27959. .28020
                 /note="AluY repeat: matches 74. .299 of consensus"
repeat_region    28028. .28269
                 /note="L1MC3 repeat: matches 6941. .7003 of consensus"
misc_feature     28146. .28221
                 /note="121 copies 2 mer gg 60% conserved"
repeat_region    28512. .28663
                 /note="tandem repeat in single clone region"
repeat_region    29643. .30184
                 /note="L2 repeat: matches 2241. .2422 of consensus"
repeat_region    30368. .30450
                 /note="LTR22 repeat: matches 1. .505 of consensus"
repeat_region    31248. .31409
                 /note="MIR repeat: matches 28. .118 of consensus"
repeat_region    31416. .31494
                 /note="L2 repeat: matches 1794. .1967 of consensus"
                 /note="MIR repeat: matches 187. .261 of consensus"
```

```
Query Match
Best Local Similarity    72.7%;    Score 108.2;   DB 9;   Length 129747;
Matches 168;   Conservative 0;   Mismatches 58;   Indels 5;   Gaps 2;

Qy      1240 GAGCTAGATGAGGAACTGGCTGGTCCCTTCCCTTCTCTTCATCCGGAATGTGTCTCTGGGT 1299
             |  |||  ||   | ||| ||   ||  |||| || |  |||  ||||  ||||  |||||
Db     25037 GTGGCCTGATACTAAAGTATCAGATGGCTTCACTATTTTCATCC-GAATGCCATCCTGGGT 25095

Qy      1300 CAGAAAACAGCCCACACAGGACCACACAGATCCCGAAGCAGCACACAGATGTGGTAAGAATA 1359
             |||||||||  ||||| |||   |||| ||||||    |||| |||||| |||||||||||
Db     25096 CAGAAAACAGCCCAGAGGGGACAGACACCAGATGTGGTGAAG 25151
```

```
Db    25152 GACATCAATGGCCTTGTGGCTGGCTGGAAAAATTCATCCAGCCTACTGACCCAGCCATGATC 25211
Qy     1420 CGACCAGTGCCTGCTGTTAAGTCCAAGGCCAAGAGTAAACAGAAGAGTGAGAATA 1470
            |  ||||  |||| ||||  ||  ||||||||  ||  ||| |  || | || | | 
Db    25212 CGCCCATTGCCCATTAAGTCCAAGGCCCAAGGCCCAAGGTTAAGCAGAGAGAGCACCACA 25262
```

```
RESULT 12
AL353694_2/c
WPCOMMENT
Sequence split into 4 fragments    LOCUS AL353694 Accession AL353694
         Fragment Name    Begin    End
         AL353694_0       1        110000
         AL353694_1       100001   210000
         AL353694_2       200001   310000
         AL353694_3       300001   356330
Continuation (3 of 4) of AL353694 from base 200001 (AL353694 Homo sapiens chromosome

Query Match
Best Local Similarity    93.8%;    Score 101.8;   DB 2;   Length 110000;
Matches 106;   Conservative 0;   Mismatches 7;   Indels 0;   Gaps 0;

Qy      616 ATTGAGCTCCTACTCCAGAGCCCAGCGTTCTGAGGTGGAGAGATGATTCGAGACATGGGT 675
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    60753 ATTGAGCTTCTACTCCAGAGCCCAGCGCCTGAGGTGGAGAGATGATTCCGAGACATGGGT 60694

Qy      676 GTGGGACAGTCAGCGGTGGAGCAGCGTGGGCTGTTGGTTGACTGCGGTGTCCCTCCAAGAA 728
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    60693 GTGGGACAGTCAGCGGTGGAGCAGCGTGGGCTGTTGGTTGACTGTGTCTGCGTGTCTCCAAGAA 60641
```

```
RESULT 13
AC099330/c
LOCUS         AC099330               163424 bp    DNA    linear    HTG 09-NOV-2001
DEFINITION    Homo sapiens chromosome 3 clone RP11-78010, WORKING DRAFT SEQUENCE,
              8 unordered pieces.
ACCESSION     AC099330 AC068701
VERSION       AC099330.1 GI:16874864
KEYWORDS      HTG; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 163424)
AUTHORS       Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE         Direct Submission
JOURNAL       Unpublished
REFERENCE     2  (bases 1 to 163424)
AUTHORS       Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE         Direct Submission
JOURNAL       Submitted (09-NOV-2001) Genome Center, University of Washington,
              Box 352145, Seattle, WA 98195, USA
COMMENT       On Nov 9, 2001 this sequence version replaced gi:8247914.
              --------    Genome Center
              Center: University of Washington Genome Center
              Center Code: UWGC
              Web site: http://www.genome.washington.edu
              Contact: uwgc@u.washington.edu
              --------    Project Information
              Drafting Center: BCM
              Center project name: chr-3
              Center clone name: RP11-78010 (bc0618)
              --------    Summary Statistics
              Sequencing vector: unknown; 42% of reads
              Sequencing vector: plasmid; 58% of reads
              Chemistry: Dye-terminator ET; 92% of reads
              Chemistry: Dye-terminator Big Dye; 8% of reads
              Assembly program: Phrap; version 0.990319
              Consensus quality: 158969 bases at least 040
              Consensus quality: 160979 bases at least 030
              Consensus quality: 162023 bases at least 020
              Insert size: 202900; 21.9% error; agarose-fp
```

```
Qy      1360 GGCTTGATGGGCTTGGAGGACGAACAACAAAAATTCATCCAGCCTAGGGACACACAACCATTATC 1419
             |  |||  |||||||||||||||||| ||| |||||||||||||||||||||||||||||||
```

```
          Insert size: 162724; sum-of-contigs
          Quality coverage: 9.0x in Q20 bases; agarose-fp
          Quality coverage: 11.2x in Q20 bases; sum-of-contigs

        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 8 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.

FEATURES             Location/Qualifiers
     source          1..163424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone_lib="RPCI human BAC library 11"
                     /clone="RP11-78O10"
     misc_feature    1..4608
            1         4608: contig of 4608 bp in length
            4609      4708: gap of unknown length
                     /note="assembly_name:Contig74"
     misc_feature    4709..10289
            4709      10289
            10290     10389: gap of unknown length
                     /note="assembly_name:Contig75"
     misc_feature    10390..17105
            10390     17105: contig of 6716 bp in length
            17106     17205: gap of unknown length
                     /note="assembly_name:Contig76"
     misc_feature    17206..45100
            17206     45100
            45101     45200: gap of unknown length
                     /note="assembly_name:Contig77"
     misc_feature    45201..65174
            45201     65174
            65175     65274: gap of unknown length
                     /note="assembly_name:Contig78"
     misc_feature    65275..90719
            65275     90719
            90720     90819: gap of unknown length
                     /note="assembly_name:Contig79"
     misc_feature    90820..119377
            90820     119377
            119378    119477: gap of unknown length
                     /note="assembly_name:Contig80"
     misc_feature    119478..163424
            119478    163424: contig of 43947 bp in length
                     /note="assembly_name:Contig81"
BASE COUNT    40950 a    40937 c    39779 g    40912 t    846 others
ORIGIN
```

```
RESULT 14
FR165K09/c
LOCUS       FR165K09    39410 bp    DNA    linear    VRT 23-JAN-2002
DEFINITION  Fugu rubripes cosmid 165K09 DNA for GRM7, TRIP, Sand, PRGFR3 genes.
ACCESSION   AJ010317
VERSION     AJ010317.1  GI:3928163
```

```
Query Match    93.8%;  Score 101.8;  DB 2;  Length 163424;
Best Local Similarity    5.2%;
Matches 106;  Conservative 0;  Mismatches 7;  Indels 0;  Gaps 0;    Pred. No. 1.4e-17;

Qy    616  ATTGAGCTCCTACTCCAGAGCCACGCGTTCTGAGGTGGAGGAGGAGATGAATTCGAGACATGGGT 675
           |||||||| |||||||||||||||||| ||||||||||||||||||||||||| ||||||||||||
Db  55015  ATTGAGCTTCTACTCCAGAGCCACGCGCCCTGAGGTGGAGGAGGAGATGAATTCCGAGACATGGGT 54956

Qy    676  GTGGGACAGTCAGCGGGTGGAGCAGCTGGCTGTGTGTACTGCGCTGTCCCTCAAGAA 728
           |||||||||||||||||||| |||||||||||||||| |||||| ||||| |||||||
Db  54955  GTGGGACAGTCAGCGGGTGGAACAGCTGGCTGTGTGTGTACTGCGTGTCTCTCAAGAA 54903
```

```
KEYWORDS    GRM7 gene; metabotropic glutamate receptor 7; plasminogen related
            growth factor receptor 3; PRGFR3 gene; Sand gene; TRAP interacting
            protein; trip gene.
SOURCE      Takifugu rubripes
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (sites)
  AUTHORS   Cottage,A., Clark,M., Hawker,K., Umrania,Y., Wheller,D., Bishop,M.
            and Elgar,G.
  TITLE     Three receptor genes for plasminogen related growth factors in the
            genome of the puffer fish Fugu rubripes
  JOURNAL   FEBS Lett. 443 (3), 370-374 (1999)
  MEDLINE   99148833
   PUBMED   10025966
REFERENCE   2  (sites)
  AUTHORS   Cottage,A.J., Edwards,Y.J. and Elgar,G.
  TITLE     SAND, a new protein family: from nucleic acid to protein structure
            and function prediction
  JOURNAL   Compar. Funct. Genom. 2, 226-235 (2001)
REFERENCE   3  (bases 1 to 39410)
  AUTHORS   Cottage,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-1998) Cottage A.J., HGMPRC, Computing, Hinxton,
            Cambridge, CB10 1SB, UK
FEATURES             Location/Qualifiers
     source          1..39410
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone="cosmid 165K09"
     mRNA            join(<261..401,496..696,1891..2030,2155..2778,2854..3165,
                     4227..>4466)
                     /gene="GRM7"
     gene            261..4466
                     /gene="GRM7"
     CDS             join(<261..401,496..696,1891..2030,2155..2778,2854..3165,
                     4227..4466)
                     /gene="GRM7"
                     /codon_start=3
                     /product="metabotropic glutamate receptor 7"
                     /protein_id="CAA09083.1"
                     /db_xref="GI:3928164"
                     /db_xref="SPTREMBL:O9PWQ0"
                     /translation="FDQYFTSRSRENRNRNIWFAREFWEDDFKCKLTRPGIKYELGRRK
                     CTGDERISRDSQYEQEGKVQFVIDAVYAMAHALHSMHTDLCPGSMGVCDKMDPVGRM
                     LQYIHGVNFNGSAGTGVMFNENGDAPGRYDIFQYQMSNISNPGYRNICGOWTNHLRLN
                     LEEMQWSGGDRKIPESVCSFPCESGERKKMVKGVPCCWHCELCDGYQYLLDEPTCDMC
                     PYDMRPLKNRTGCRPTPIKLEWSSPWAIIPVFLALIGLIAITGVIATFIRNDTPIV
                     RASGRELSYVLLTGIFLIYLITFLMIAEPSVAVCAPRRLFLGLGMCISYSAMLTKTNR
                     IYRIFEQGKKSVTPPKFISPTSQLIITFILISVQLLGVFIWFGVMPPHTIIDYEEQKP
                     PNPEFARGVLKCDMSDLSLICLSYSLVLMITCTVYAIKSRGVPETFNEAKPIGFTMY
                     TTCIVWLAFVPIFFGTAQSTEKMFIQTTTLVSMSLSATVSLGMLYIPKVYVIIFHPE
                     QNVQKRKRSEFKAVVQAATVSTRLSQKSSDKQNGESKIEPDRSQ"
     exon            <261..401
                     /gene="GRM7"
                     /number=1
     intron          402..495
                     /gene="GRM7"
                     /number=1
     exon            496..696
                     /gene="GRM7"
                     /number=2
     intron          697..1890
                     /gene="GRM7"
                     /number=2
     exon            1891..2030
                     /gene="GRM7"
                     /number=3
     intron          2031..2154
                     /gene="GRM7"
                     /number=3
```

```
exon        2155..2778
            /gene="GRM7"
            /number=4
intron      2779..2853
            /gene="GRM7"
            /number=4
exon        2854..3165
            /gene="GRM7"
            /number=5
intron      3166..4226
            /gene="GRM7"
            /number=5
exon        4227..>4466
            /gene="GRM7"
            /number=6
mRNA        complement(join(<4660..4801,6176..6226,6468..6521,
            6616..6743,6799..6931,7017..7106,7205..7292,7382..7495,
            7643..7737,7816..7943,8029..8068,8194..8277,8358..8415,
            8496..>8593))
            /gene="TRIP"
CDS         complement(join(4660..4801,6176..6226,6468..6521,
            6616..6743,6799..6931,7017..7106,7205..7292,7382..7495,
            7643..7737,7816..7943,8029..8068,8194..8277,8358..8415,
            8496..8593))
            /gene="TRIP"
            /codon_start=1
            /product="TRAF interacting protein"
            /protein_id="CAA09084.1"
            /db_xref="GI:3928165"
            /db_xref="SPTREMBL:Q9YGN2"
            /translation="MPIRAQCTICSDFPDHSRDVAAIHCGHTFHHECLVRWFQTAPTK
            TCPQCRKQVSTRHIISRLYFDVGLDDSSVGDPESLQNELDRKVNFSSKERDWREKQK
            AMDLMEFVELQRKALENLQKDVWEKEKLCSALRTQMTYLESQHNDTRAAKEEVRRLR
            IKNKTFESLDVVLQGQRAEVESMITDMGIGQOAAVEQLSIYCISLKKEYDNLKGGLKSS
            NEMCEKLKREVLITSNNKLHKALLELTKARDDMKSLQNDLTNAEKEISSLKKKVEFLQE
            ALSTPTRTNEALCRRLFEERCGAAPPQNSRVSKASCLHPPSGNEDIDLNMTYDVTTPDD
            VRKRPKKVPSKMHLDSVVILRTKYKNLVSNQGKFVQSGYDGLGGRTKFIQPRTIHA
            VRSEASKPENSGTSLDLYTSRKEIDFSGGELKNSLFLGCGK"
gene        complement(4660..8593)
            /gene="TRIP"
exon        complement(4660..4801)
            /gene="TRIP"
            /number=14
intron      complement(4800..6177)
            /gene="TRIP"
            /number=13
exon        complement(6176..6226)
            /gene="TRIP"
            /number=13
intron      complement(6225..6469)
            /gene="TRIP"
            /number=12
exon        complement(6468..6521)
            /gene="TRIP"
            /number=12
intron      complement(6520..6617)
            /gene="TRIP"
            /number=11
exon        complement(6616..6743)
            /gene="TRIP"
            /number=11
intron      complement(6742..6800)
            /gene="TRIP"
            /number=10
exon        complement(6799..6931)
            /gene="TRIP"
            /number=10
intron      complement(6930..7018)
            /gene="TRIP"
            /number=9
exon        complement(7017..7106)
            /gene="TRIP"
            /number=9

intron      complement(7105..7206)
            /gene="TRIP"
            /number=8
exon        complement(7205..7292)
            /gene="TRIP"
            /number=8
intron      complement(7291..7383)
            /gene="TRIP"
            /number=7
exon        complement(7382..7495)
            /gene="TRIP"
            /number=7
intron      complement(7494..7644)
            /gene="TRIP"
            /number=6
exon        complement(7643..7737)
            /gene="TRIP"
            /number=6
intron      complement(7736..7817)
            /gene="TRIP"
            /number=5
exon        complement(7816..7943)
            /gene="TRIP"
            /number=5
intron      complement(7942..8030)
            /gene="TRIP"
            /number=4
exon        complement(8029..8068)
            /gene="TRIP"
            /number=4
intron      complement(8067..8195)
            /gene="TRIP"
            /number=3
exon        complement(8194..8277)
            /gene="TRIP"
            /number=3
intron      complement(8276..8359)
            /gene="TRIP"
            /number=2
exon        complement(8358..8415)
            /gene="TRIP"
            /number=2
intron      complement(8414..8497)
            /gene="TRIP"
            /number=1
exon        complement(8496..>8593)
            /gene="TRIP"
            /number=1
exon        <9737..9860
            /gene="Sand"
```

```
RESULT 15
I66494/c
LOCUS        I66494        7218 bp    DNA        linear    PAT 28-DEC-1997
DEFINITION   Sequence 14 from patent US 5670367.
ACCESSION    I66494
VERSION      I66494.1  GI:2724471
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unknown.

Query Match       3.7%;  Score 72.4;  DB 5;  Length 39410;
Best Local Similarity   77.2%;  Pred. No. 3.5e-09;
Matches  88;  Conservative  0;  Mismatches  26;  Indels  0;  Gaps  0;

Qy     112 ATGCCTATCCTTCTCTGTGCACTATCTGCTCCGGACTTCTTCGATCACTCCCGTGACGTG 171
           |||||||||||| | |||||| |||| ||| ||||| ||||||||||||| || || ||
Db    8593 ATGCCTATCCGAGCACAGTCGCACAATTTGTTCCGACTTCTTCGATCACACTCCAGAGATGTT 8534

Qy     172 GCTGCCATCCACTGTGGCCACACTTTTCATCTGCAATGCCTAATCCAGTCGGTTT 225
           || |||||||||||| ||||| |||||| || || ||| |||||||||| |||||
Db    8533 GCAGCCATCCACTGCGGACACACTTTTCCACCATGAATGGTTCATCCAGTTTGTT 8480
```

RESULT 16

LOCUS   AF410847   2065 bp   DNA   linear   VRL 20-SEP-2001
DEFINITION Ovine herpesvirus 2 latency associated antigen (orf73) gene,
            complete cds.
ACCESSION   AF410847
VERSION   AF410847.1 GI:15705839
KEYWORDS   .
SOURCE   Ovine herpesvirus 2.
  ORGANISM Ovine herpesvirus 2.
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae.
REFERENCE   1 (bases 1 to 2065)
  AUTHORS   Coulter,L.J. and Reid,H.W.
  TITLE   Isolation and expression of three open reading frames (ORFs) from
            ovine herpesvirus 2
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2065)
  AUTHORS   Coulter,L.J. and Reid,H.W.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-AUG-2001) Virology, Moredun Research Institute,
            Pentlands Science Park, Bush Loan, Penicuik, Midlothian EH26 0PZ,
            UK
FEATURES   Location/Qualifiers
     source   1..2065
                     /organism="Ovine herpesvirus 2"
                     /virion
                     /db_xref="taxon:10398"

     gene   complement(63..1550)
                     /gene="orf73"
     CDS   complement(63..1550)
                     /gene="orf73"
                     /note="orf73"
                     /note="LANA; similar to HVS and HHV-8 Orf73"
                     /codon_start=1
                     /product="latency associated antigen"
                     /protein_id="AAL05844.1"
                     /db_xref="GI:15705840"
                     /translation="MVLLRSGTSTSTDGDGEDGRRPGPKRPVTTGKGKGEVEGPGGPGE
                     GPEGEGPGEGPGEGPEGEGPGEGPGEGPGEGPGEGPEGEGPGEGPEGEGPGEGPGEGPGE
                     GEGPGGEVEGPGEGPGEGPGEGPGEGPGEGPEGEGPGEGPGEGPEGEGPGEGPEGEGPGE
                     VEGPGEGPEGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPGEGPEGEGPGEGPEGEGPG
                     EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
                     EGPGGEGGEGPGEGEGGSSGGPPVRGRGRPGRGEGGSSGGPVVRGRSRVRALNPNAPHSN
                     RKKLAPIWDPTLKEATYSLHLNCTSKDPVVRVSRSVRALNPNAPHSNIFTGGMWTFV
                     IYGNDKEAVESLFQFLLQDAMNPDQAGAVNISTGPLTPSLPFNQQ"

BASE COUNT   351 a   684 c   303 g   727 t
ORIGIN

Query Match   2.9%; Score 57.6; DB 14; Length 2065;
Best Local Similarity   45.7%; Pred. No. 5.4e-05;
Matches 201; Conservative   0; Mismatches 239; Indels   0; Gaps   0;

Qy   490 AAGGCAGAGATGCTGTGTTCCACCCTGAAAAAACAGATGAAGAGTTCCTGGACCAGCGCAG 549
          |||| || || || ||| || ||| | ||| || || || || ||||||| |||
Db   1162 AAGGAGAAGGACCTGGAGGAGAAGTAGAAGGTCCTGGAGGAGAAGGACCTGGAG 1103

Qy   550 GATGAGACCAAAACAAGCTCGGGAGGAGGCCCACCGACTCAAGTGCAAGATGAAAACCATG 609
          || | || || | |||||| || || || |||| |||||| || ||| ||| ||
Db   1102 GAGACTAGAAGAGGTCCTGGAGGAGAAGGACCTGGAGGAGAAGGACCTG 1043

Qy   610 GAGCAAAATTGAGCTCCTACTCCAGAGCCACCAGGGAAACTGGCTGACAGGTTGAAG 669
          ||| || |||| || | |||||||| || ||||| |||||||| ||||||||
Db   1042 GAGGAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGAC 983

Qy   670 ATGGGTGTGGGACAGTCAGCCGGTGGAGCACTTACAAAGTGCTGACCAGGAGATCACGAGC 729
          |||| || |||||||||||||||||||| | |||| |||||| | |||| ||||||||
Db   982 CTGGACGAGAAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGACCTG 923

Qy   730 GAGTATGAGAATCTGAAGGAAGAGTCTGGAAAGGCCCACAGGGAAACTGGCTGACAGGTTGAAG 789
          ||| || || || || ||| |||| ||| || || || ||||| ||||| |||||| |||
Db   802 GAGAAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGAC 743

Qy   790 AAGGATTGGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACACTGAGCTGGATCAGCGC 849
          |||| | || || || ||||| ||| ||| ||| | |||||| ||| | |||
Db   862 AAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGAC 803

Qy   850 AAGTTAGAACTGAGGTCAGCCCAGAAGAGGACTTACAAAGTGCTGACCAGGAGATCACGAGC 909
          ||| | || |||||| | ||||||||||||| | |||| ||||| | |||| ||||||
Db   802 GAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGACCTGGAGGAGAAGGAC 743

Qy   910 CTAAGAAAGAAGTCTGATGA 929
          | ||| ||||| ||||||
Db   742 AAGAACAGGAAGAAGAAGGA 723

RESULT 17

LOCUS   AC067854   203984 bp   DNA   linear   HTG 04-JAN-2002
DEFINITION Homo sapiens chromosome 8 clone RP11-513H8 map 8, *** SEQUENCING IN
            PROGRESS ***, 1 ordered pieces.
ACCESSION   AC067854
VERSION   AC067854.13 GI:18056699
KEYWORDS   HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 203984)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```
TITLE     Homo sapiens chromosome 8, clone RP11-513H8
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 203984)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
          Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
          Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
          Campoplano,A., Castle,A., Colangelo,M., Collins,S.,
          Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
          Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
          Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
          Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
          Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
          Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
          McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
          Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
          Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
          Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
          Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
          Tesfaye,S., Theodore,J., Tirelli,A., Travers,M., Trigilio,J.,
          Vassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
          Young,G., Zainoun,J., Zimmer,A. and Zody,M.
          Direct Submission
JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jan 4, 2002 this sequence version replaced gi:18034761.
          All repeats were identified using RepeatMasker:
          Smit,A.F.A. & Green,P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ----------------  Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          ----------------  Project Information
              Center project name: L9308
              Center clone name: 513_H_8
          ----------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *    1  203984: contig of 203984 bp in length.
FEATURES             Location/Qualifiers
     source          1..203984
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-513H8"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    61400 a   39486 c   39964 g   63134 t
ORIGIN
```

```
Qy    314 AGGAAGAGGAAGTCTTGGGATCGGAGAATTCTTAAAGAATGAACTGGACACGCGTCAAAG 373
          |||| |||||||||||||||||||||||| |||| ||||| ||||||||||||||||| |
Db  16618 AGGAGGAGGAAGAGGAGGAGGAGGAGGAAGAGAAGGAGGAGGAGGAGGAGGAGGAAAGG 16559

Qy    374 CTCAGCTTTCCCAGAAAGACAGGGAGGAGAAACGGGACACAGCCACGGCCATTATCGACACTCTAC 433
          || |||||||                |||| ||| |||
Db  16558 AGGAGGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGGAGGAGAAGGAGG 16499
```

```
Query Match          2.8%;  Score 55.2;  DB 2;  Length 203984;
Best Local Similarity   43.4%;  Pred. No. 0.00033;
Matches  255;  Conservative   0;  Mismatches 333;  Indels   0;  Gaps   0;
```

```
Qy    434 GGGACACCCTGGAAGAACGCAATGCTACCGTGGAGTCCCTACAGAACGCGTGGAAG 493
          ||||||                |||||||||||||||||| |  || |||| ||||
Db  16698 AGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGGAGATG 16439

Qy    494 CAGAGATGCTGTGTTCCACCCTGAAAAACAGATGAAGTTCCTGGACGAGGAAGATG 553
          ||||| |              ||| | || | | || | ||| ||| | |||| |
Db  16438 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAGG 16379

Qy    554 AGACCAAAACAAGCTCGGGAGGAGCCCCACCCACTCAAGTGCAAAGATGAAAAACCATGGAGC 613
          |                ||||||||||||| ||||| ||  |||           |||| |
Db  16378 AGGAGGAGGAGGAAGAAGGAGGAGGAGGAGGAGGAAAAGGAGGAGGAGGAGGAAGGAGGAAGG 16319

Qy    614 AAATTGAGCTCCTACTCCAGACCCACCGTTCTGAGGTGGAGGAGGATGATTCGGAGGACATGG 673
          |||                  ||| | |||| || | ||||||  |   || | ||| |||
Db  16318 AGGAAGAGGAGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAAGGAGGAGGAGGAGGAGGAGAAGG 16259

Qy    674 GTGTGGGCAGTCAGCCGTGGAGCAGTGGCTGTTGTGGGAGGAGGAGGAGGAAGAGT 733
          ||  ||  |||      |                |  || | || |||||||||||
Db  16258 AGAAGGAGGAGGAGGAGGAGAAGGAGGAGGAGGAGGAGAAGGAGGAGGAGAGGAGA 16199

Qy    734 ATGAGAATCTGAAGGAAGCTTCGGAAGGCCACAGGGGAACCTGGCTGACAGGTTGAACAAG 793
          |       |||  |  | ||   |||||  |||   || |||| ||| ||||  ||| |||
Db  16258 AGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGG 16199

Qy    794 ATTTGGTGTCCTCTAGGACGCAAGTTGAAGACTCTCAACACTGAGCGTGGATCAGGCCAAGT 853
          |       |||||         ||  |  || | |||     || | ||| ||| ||| |
Db  16198 AGGAGAAGGAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAGGAGGAGGAGGAGGAGG 16139

Qy    854 TAGAACTGAGGTCAGCCCAGAAGGACTTACAAAGTGCTGACCAGGAGA 901
          |      ||| | || |||| ||            |  || ||||| |
Db  16138 AGAAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 16079

Qy        ------
Db  16078 AGAAGGAGGAGGAGGAAGGGGAGGAGGGAGGAGGAGGAGGAGGAAGGAGGA 16031
```

```
RESULT 18
LMFLCHR34_16
WPCOMMENT
          Sequence split into 18 fragments  LOCUS LMFLCHR34  Accession AL499623
```

Fragment Name	Begin	End
LMFLCHR34_00	1	110000
LMFLCHR34_01	100001	210000
LMFLCHR34_02	200001	310000
LMFLCHR34_03	300001	410000
LMFLCHR34_04	400001	510000
LMFLCHR34_05	500001	610000
LMFLCHR34_06	600001	710000
LMFLCHR34_07	700001	810000
LMFLCHR34_08	800001	910000
LMFLCHR34_09	900001	1010000
LMFLCHR34_10	1000001	1110000
LMFLCHR34_11	1100001	1210000
LMFLCHR34_12	1200001	1310000
LMFLCHR34_13	1300001	1410000
LMFLCHR34_14	1400001	1510000
LMFLCHR34_15	1500001	1610000
LMFLCHR34_16	1600001	1710000
LMFLCHR34_17	1700001	1720777

```
Continuation (17 of 18) of LMFLCHR34 from base 1600001 (AL499623 Leishmania major chr
```

```
Query Match          2.7%;  Score 54.2;  DB 2;  Length 110000;
Best Local Similarity   46.7%;  Pred. No. 0.00063;
Matches  216;  Conservative   0;  Mismatches 238;  Indels   0;  Gaps   1;
```

```
Qy    491 AGGCAGAGATGCTGTGTTGTGTTCCACCCTGAAAAACAGATGAAGTTCCTGGACGAGGCGCAGG 550
          |||| ||||| ||||||| |||||||||||| | |  | || || | ||| |||||| |||
Db  46336 AGGCCGAGAGCCTGCCTGCCTGCCGAGCTCGGAGAGGCGGCCACCGACTCAAGTGCAGAGCCTG 46395

Qy    551 ATGAGACCAAACAAGCTCGGGAGGAGCCCCACCGAGGCGCCGAGAGCTCAAGTGCAAGAGGAAAACCATGG 610
          ||||| |||   ||||| |||                |   ||||| ||| | | |||||| |||||
Db  46396 CCGACCTCGAGCAAGCGGGGAGGAGGAGGAGGAGGAGGAGCCGAAAGGCAGAAGCAGAGGCCCGA 46446

Qy    611 AGCAAATTGAGCTCCTACTCCAGAGCCACGCGTTCTGAGGTGGAGGAGATTCGAGACA 670
          |                  | |||||| |  |                   || |||
Db  46446 ------------------CAACG
```

```
Db  46447  AGCAGCTGGCACGCCGAGCTGAACAGGGCCGCAGGAGGAGGCTGAGAAGCTGGCTGCCGAGC
           |||      |||    ||||||  |||| ||| |   |||| ||||  |||| |||
Qy    671  TGGGTGTGGGACAGTCACGGGTGAGCACCTGGCTGTTGTTGGTGTTCCCTCAAGAAAG 730

Db  46507  TGGAGAAGGCGCAGGAGGAGGCTGAGAAGCTGGCTGCCGACCTCGAGAAGCGGGAGGAGG
           |||    ||| ||||  |||||  ||||||||||||
Qy    731  AGTATGAGAATCTGAAGGAAGCTCGGAGAAGGCCACAGGGGAACTGGCTGACAGGGTTGAAGA 790

Db  46567  AGGCCGAGAGGACAGAAGGCGCGACAAACGACCGGCGTGGCCCCCGACCTGAAGAGGGGCGCAGG
           ||||  ||||     |||| || ||   ||| ||| ||||||   |||||| |
Qy    791  AGGATTTGGTTGGTCCTCTAGGAGCAAGTTGAAGACTTCAACACTGAGCTGGATCAGGGCCA 850

Db  46627  AGGAAGGCCGAGAGGCTTGCTGCCCGAGTTGGACACGAGGCCGGAGGCCGAGAGGCTGG 46686
           |||||    ||||    || |||        |||       |||  |||
Qy    851  AGGTTAGAACTGAGGTCAGCCCACAGAAGGACTTACAAAAGTGCTGACCAGGAGATCACGAGCC 910

Db  46687  CTGCCGAGCTGGAGAGGGGCGCCAGGAGGCCGAGAGGCTGGCCCCGAGAGCTGAACAGGG 46746
           |||  ||||| ||||  |||  |||||| || |||  ||||||||  |
Qy    911  TAAGAAAGAAGTCTGATGATCCTCCAGGGAACCTTGAGCCTGC 953

Db  46747  CGCAGGAGGAGGCTGAGAAGCTGGCTGCCAACCTCGAGAAGGCC 46789
```

```
RESULT 19
LOCUS        AC098712              207683 bp    DNA    linear    HTG 14-NOV-2001
DEFINITION   Mus musculus chromosome UNK clone RP23-2A21, WORKING DRAFT
             SEQUENCE, 5 unordered pieces.
ACCESSION    AC098712
VERSION      AC098712.2  GI:16924164
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 207683)
  AUTHORS    McPherson,J.D. and Waterston,R.H.
  TITLE      The sequence of Mus musculus clone
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 207683)
  AUTHORS    McPherson,J.D. and Waterston,R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park
             Parkway, St. Louis, MO 63108, USA
COMMENT      On Nov 14, 2001 this sequence version replaced gi:16554409.
             ---------------- Genome Center --------------------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc/index.shtml
             Contact: submissions@watson.wustl.edu
             ---------------- Project Information ----------------
             Center project name: M_BA0002A21
             ---------------------------------------------------
             ---------------- Summary Statistics ----------------
             Sequencing vector: M13; %
             Sequencing vector: plasmid; %
             Chemistry: Dye-primer ET; % of reads
             Chemistry: Dye-terminator Big Dye; % of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 205930 bases at least Q40
             Consensus quality: 206594 bases at least Q30
             Consensus quality: 207083 bases at least Q20
             Insert size: -; agarose-fp
             Insert size: 21954; sum-of-contigs
             Quality coverage: 25.41 in Q20 bases; agarose-fp
             Quality coverage: 24.42 in Q20 bases; sum-of-contigs
             --------------------------------------------------.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 5 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
             1              1240: contig of 1240 bp in length
             1241          1340: gap of unknown length
             1341         14229: contig of 1889 bp in length
             14230        14329: gap of unknown length
             14330        39801: contig of 25472 bp in length
             39802        39901: gap of unknown length
             39902        71680: contig of 31779 bp in length
             71681        71780: gap of unknown length
             71781       207683: contig of 135903 bp in length.
FEATURES             Location/Qualifiers
     source          1..207683
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-2A21"
     misc_feature    1..1240
                     /note="assembly_name:Contig14"
     misc_feature    1341..14229
                     /note="assembly_name:Contig38"
     misc_feature    14330..39801
                     /note="assembly_name:Contig37"
     misc_feature    39902..71680
                     /note="assembly_name:Contig39"
     misc_feature    71781..207683
                     /note="assembly_name:Contig40"
BASE COUNT    62336 a  44191 c  44243 g  56460 t  453 others
ORIGIN
```

```
Query Match    2.78;    Score 53.8;  DB 2;    Length 207683;
Best Local Similarity   46.4%;    Pred. No. 0.00084;
Matches  175;  Conservative  0;  Mismatches 202;  Indels  0;  Gaps  0;

Qy    524  AGATGAAGTTCCTGGACACCAGGGCCAGGATGAGACCAAAACAAGCTCGGGAGGAGGCCCACC 583
           ||| |||| || |||||  |||| |||| |||   ||  |||  |||||  ||| |||| ||
Db 186178  AGCAGGAGGAGGACGAGGCCAGGAGGAGGAGGAGGAGAGCAGGAGGAGGAGGAGGAGGAGG 186237

Qy    584  GACTCAAGTCAAGATGAAGAAACCATGAGCAAAATTGAGCTCCTACTCCAGACCCAGCGTT 643
           || |  || |      ||| ||   |||| |||| ||||| ||    |||    |||
Db 186238  AGCAGGAGGACCAGGAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGCAGGAGG 186297

Qy    644  CTGAGGTTGGAGGAGGATGATGATTCGAGACATGGGTGTCGTGGGACGCTCAGCGGTGGAGCAGCTGG 703
           || |||| |||||  |||   ||   ||||  ||   ||  |||| ||||  |||||| |
Db 186298  AGGAGGAGCAGGAGCAGGAGGAGGACCAGGAGGAAGAGGAAGAGGAGCAGGAGG 186357

Qy    704  CTGTGTACTGCGGTGTCCCTCAAGAAAGAGAGTATGAGAAATCTGAAGGAAGCTCGGAAGGCCA 763
           || ||| | |||   |||  ||| ||||   || ||     || |||  |||  || || ||
Db 186358  AGGAGCAGGAGGAGGAGCAGGAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGGAGG 186417

Qy    764  CAGGGGAACTGGCTGACAGGGTTGAAGAAGGATTTGGTGTCCTTCTAGGAGCAAGTTGAAGA 823
           || ||||  || |||| ||||  |||  ||   || || ||   ||| ||||  ||
Db 186418  AGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGAAGGAGCAGGAGGAGCAGGAGGAGCAGGAGG 186477

Qy    824  CTCTCAACACTGAGCTGGATCAGCCCAAGTTAGAACTGAGGTTCAGCCCAGAGAAGGACTTAC 883
           || | ||  |  |||  |||| ||  |||  |||  |||||||| |||  |||||||
Db 186478  AGCAGGAGGAAGGAGCAGGAGCAGCAGGAGGAGCAGGAGGAGGAGCAGGAGGAGCTGAGGGGAGC 186537

Qy    884  AAAAGTCCTGACCAGGAG 900
           |  || || ||||||||
Db 186538  CGGAGGAGGAGCAGGAG 186554
```

```
RESULT 20
LOCUS        AC090495              216191 bp    DNA    linear    HTG 27-OCT-2001
DEFINITION   Mus musculus clone RP23-60B7, WORKING DRAFT SEQUENCE, 23 unordered
             pieces.
ACCESSION    AC090495
VERSION      AC090495.4  GI:16506394
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 216191)
  AUTHORS   McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
            Bastide,M., Kuit,K., King,L., Kirchoff,K.A., Miller,B.,
            Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Santos,L.,
            Shah,R.S., Spiegel,L.A., Palmer,L., Yang,C. and Zutavern,T.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216191)
  AUTHORS   McCombie.W.R.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Direct Submission
            Submitted (24-FEB-2001) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     On Oct 27, 2001 this sequence version replaced gi:16117559.
            -----------------------------------------
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact: mccombie@cshl.org
            -----------------------------------------  Project Information
            Center project name: RP23-60B7
            Center clone name: RP23-60B7
            -----------------------------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 23 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *         1:    22188: contig of 22188 bp in length
          *     22189:    22512: gap of unknown length
          *     22513:    41019: contig of 18507 bp in length
          *     41020:    41343: gap of unknown length
          *     41344:    59365: contig of 18022 bp in length
          *     59366:    59689: gap of unknown length
          *     59690:    73640: contig of 13951 bp in length
          *     73641:    73964: gap of unknown length
          *     73965:    87376: contig of 13412 bp in length
          *     87377:    87700: gap of unknown length
          *     87701:   101024: contig of 13324 bp in length
          *    101025:   101348: gap of unknown length
          *    101349:   112980: contig of 11632 bp in length
          *    112981:   113304: gap of unknown length
          *    113305:   124819: contig of 11515 bp in length
          *    124820:   125143: gap of unknown length
          *    125144:   136091: contig of 10948 bp in length
          *    136092:   136415: gap of unknown length
          *    136416:   144125: contig of 7710 bp in length
          *    144126:   144449: gap of unknown length
          *    144450:   152005: contig of 7556 bp in length
          *    152006:   152329: gap of unknown length
          *    152330:   159587: contig of 7258 bp in length
          *    159588:   159911: gap of unknown length
          *    159912:   167134: contig of 7223 bp in length
          *    167135:   167458: gap of unknown length
          *    167459:   173757: contig of 6299 bp in length
          *    173758:   174081: gap of unknown length
          *    174082:   179850: contig of 5769 bp in length
          *    179851:   180074: gap of unknown length
          *    180075:   185555: contig of 5381 bp in length
          *    185556:   185879: gap of unknown length
          *    185880:   191144: contig of 5265 bp in length
          *    191145:   191467: gap of unknown length
          *    191468:   196412: contig of 4945 bp in length
          *    196413:   196735: gap of unknown length
          *    196736:   201258: contig of 4523 bp in length
          *    201259:   201581: gap of unknown length
          *    201582:   205749: contig of 4168 bp in length
          *    205750:   206072: gap of unknown length
          *    206073:   209959: contig of 3887 bp in length
          *    209960:   210282: gap of unknown length
          *    210283:   213540: contig of 3258 bp in length
          *    213541:   213863: gap of unknown length
          *    213864:   216191: contig of 2328 bp in length.
FEATURES             Location/Qualifiers
     source          1..216191
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-60B7"
BASE COUNT    60680 a   45448 c   44115 g   58724 t    7224 others
ORIGIN
```

```
Query Match       2.7%;  Score 53.2;  DB 2;  Length 216191;
Best Local Similarity   43.0%;  Pred. No. 0.0013;
Matches 259;  Conservative    0;  Mismatches 343;  Indels    0;  Gaps    0;

Qy      313 CAGGAGACAGGAGAGAATGTCTTGGATGCAGAATTCTTAAAGAATGAACTGGACAAGCGTCAAA 372
            |||| ||||||| ||| || | |||  |||  ||  ||||||||||| | |||||| |||| |||
Db   109137 CAGGACCAGGAGCCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAG 109078

Qy      373 GCTCAGCTTTCCCAGAAAGACAGGAGAAAACGGGACAGCCAGCCCATTATCGACACTCTA 432
            |||  |  |  ||||  |  ||||| |||| ||||||||||||||||||  |  ||| |||
Db   109017 GAGCAGGAGCTGGAGCAGGAGGAGTAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGCAG 108958

Qy      433 CGGGAGCACCCTGGAAGAAAACGCAAATGCTTACCCTGGAGTCCCTACAGAACGCCTTAAACAAG 492
            | |||| |   |||| | ||| ||  || | |||| || || || || || |||||    ||||
Db   109077 CAGGAGCAGGAGCCAGGAGCAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAG 109018

Qy      493 GCAGAGATGCTGTGTTCCACCCTGAAAAAACAGATGAAGTTCCTGGAGCACGCGGCCAGGAT 552
            ||| ||   ||| || ||  || ||||| |  || |||||  || |||||   ||  ||||
Db   109017 GAGCAGGAGCTGGAGCAGGAGGAGTAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGCAG 108958

Qy      553 GAGACCAAAACAAGCTCGGGAGGAGGGCCCACCCGACTCAAGTGCCAAGATGAAAAACCATGGAG 612
            ||||| ||  ||||| |   ||||| | | ||| ||||| |||   || || ||| |||  |||
Db   108897 GAGCAGGAGCTGGAGCAGGAGGAGTAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGCAG 108838

Qy      613 CAAATTGAGCTCCTACTTCCAAGACCAGCCGGTTCTGAGGTGGAGGAGGATGATTCGAGACATG 672
            ||  | |||||   | |  ||||  |||| ||   ||||| || |||  ||||  |||||  |||
Db   109017 GAGCAGGAGCTGGAGCAGGAGGAGTAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGCAG 108958

Qy      673 GGTGTGGGACAGTCACGCGGTGGAGCAGCTGGCTGTCGTGGTACTCCGTGTGTCCCTCAAGGAAAGAG 732
            || |||||   |||   ||||| ||||||| | || |||  |||  ||||| | ||  || |||||||
Db   108777 GAGGAGCAGCAGGAGCTGGACCAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAG 108718

Qy      733 TATGAGAATCTGAAGGAAGCTCGGAAGCTTCGGCTTGACAGGTTGAAGAAG 792
            || |||||| ||  ||||| |   |||  |  |||  |||||||||||||||
Db   108717 GAGCAGGAGCTGGAGCAGGAGGAGTAGGAGGAGCAGGAGCAGGAGGAGCTG 108658

Qy      793 GATTTGGTTGTGTCCTCTAGGAGAAAGTTGAAAGACTCTCAAACACTGAGTGAGGTCAGGCCAAG 852
            |||  |||| ||| | ||| ||||||| |||| | ||| |  |||  |||  | |||||||||||
Db   108657 GAGCTGAAGCAGGAGGAGCAGGAGGAGCAGGAGCTGGAGCTGAAGCAG 108598

Qy      853 TTAGAACTGAGGTCAGCCCCAGAAAGGACTTACAAAGTGCTGACCAGGATCACGACGACCTA 912
            ||  ||||| || || || ||||||||  |||| || |||||| |||||  |  |||||||
Db   108597 GAGGAGAAGCAGGAGCAGGAGCTGGAGCAGGAGGAGCAGGAGGAGAGTGTTTA 108538

Qy      913 AG 914
            ||
Db   108537 AG 108536
```

```
RESULT 21
AF305694
LOCUS       AF305694                3127 bp    DNA     linear   VRL 30-MAR-2001
DEFINITION  Kaposi's sarcoma-associated herpesvirus latent nuclear antigen
            gene, partial cds.
ACCESSION   AF305694
```

```
VERSION     AF305694.1  GI:11037007
KEYWORDS    .
SOURCE      Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.
  ORGANISM  Human herpesvirus 8
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1  (bases 1 to 3127)
  AUTHORS   Piolot,T., Tramier,M., Coppey,M., Nicolas,J.C. and Marechal,V.
  TITLE     Close but distinct regions of human herpesvirus 8
            latency-associated nuclear antigen 1 are responsible for nuclear
            targeting and binding to human mitotic chromosomes
  JOURNAL   J. Virol. 75 (8), 3948-3959 (2001)
  MEDLINE   21165304
   PUBMED   11264383
REFERENCE   2  (bases 1 to 3127)
  AUTHORS   Piolot,T., Marechal,V. and Nicolas,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2000) Service de Microbiologie, Hopital
            Rothschild, 33 boulevard de Picpus, Paris 75571, France
FEATURES             Location/Qualifiers
     source          1..3127
                     /organism="Human herpesvirus 8"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:37296"
                     /cell_line="BBG-1"
                     /cell_type="B-lymphoma"
     primer_bind     1..24
                     19..>3127
     CDS             1..>3127
                     /codon_start=1
                     /note="LNA1; LANA; ORF73-like"
                     /product="latent nuclear antigen"
                     /protein_id="AAG27458.1"
                     /db_xref="GI:11037008"
                     /translation="MAPGMRLRSGRSTGAPLIRGSCRKRNRSPERCDLGDDLHLQPR
                     RKIHAVDSVDGRECGPHTLPIPGSPTVFTSGLPAFVSSPTLPVASIPSPAPATPLPPPA
                     LLPVTTSSSPIPSHPVSPGTTDTHSPSSPALPPTQSPESSQRPPLSSPTGRPDSSTP
                     MRPPSQQTTPPHSPTTPEPPEPSKSSPDSLAPSTLRSLRKRLSSPGGPSTLNPICQ
                     SPPVSPRCDFANRSVYPPWATESPIYVGSSSSDGDTPRQPPTSPISICSSSSPSEGSS
                     GDTAMLVLLAEIABEASKNEKCESENNQAGEDNGDNEISKESQVKDKDNDNKDDEEE
                     QEFDEDEDEEDDEDEDDEDDEDEDDEDDEDDEDDEDDEDEDDEDDEDDEEDEEE
                     DEEDEEEDEEDDEEDEEDDEDEEDDEDEEDDEDEDEEDKKEDDEDGDGNKTLSIOS
                     SOQQOEEPQOQOEPQOQOEPQOQOEPQOQOEPLOEPOOQOEPOOQOEPQOQOEPLQ
                     EPOOQOEPOOQOEPQOQOEPOOQOEPOOQOEPQOQOEPQOQOEPQOQOEPQOQOEP
                     ELEOEOELEEOQOEOOOEOEEOQOELEEOQELEOEOELEEOQELEOEOELEEOOEL
                     OQEOELEEQOELEEOQOEOEEOQELEOQOELEOQOELEEOQELEOEOELEEOOEOQ
                     VEQQEOELEEVEEOOEQQELEEVEEQOELEEVEEOOELEEVEEOOELEEVEEOEOOG
                     VEQQEOETVEEPIILHGSSSEDEMEVDYPVVSTHEQIASSPPGONTPDDDPOPSRE
                     YRVLRTSPPHRPGVRMRVPVTHPKHPRYQOPPYPVLRKPRGHSRPIQIDPAKARPQHIFYRR
                     FLGKDGRRDPKCQWKFAVIFWGNDPYGLKKLSQAFOFGGVKAGPVSCLPHPGPQSPI
                     TYCVYVYCQNKDTSKKVQMARLAWEASHPLAGNLQSSIVKFKKPLPLTQPGENQGPGD
                     SPDEMT"
ORIGIN
```

```
BASE COUNT      941 a      744 c      1009 g      433 t
     primer_bind     complement(3104..3127)
```

```
Query Match             2.6%;  Score 51.6;  DB 14;  Length 3127;
Best Local Similarity  45.7%;  Pred. No. 0.0029;
Matches  180;  Conservative    0;  Mismatches 214;  Indels   0;  Gaps   0;

Qy    538 GAGCACGCGGCAGGAtGAGACCAAACAAGCTCGGGAGGAGCCCACCGACTCAAGTGCAAG 597
          |||||||||||||||| |||||||||||||  || |||||| |  |  ||||||| || 
Db   1699 GAGCGCACCAGGAtGAGCAGCACCAGGATGAGCAGCACCAGGATGAGCAGCACCAGCAG 1758

Qy    598 ATGAAAAACCATGGACAACAAATTGAGCTCCTCTACTCCAGAGCCAGTGGCAGGAG 657
          ||  |||| || ||  | |||  |||| |||||| || || ||||| |||||||| 
Db   1759 GATGAGCAGCACCAGGATGAACAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAG 1818

Qy    658 ATGAGTTCGAGACAtCAGGGGTTGGGACAGTGCGGTGGACAGAGCAGTTAGAGGAGG 717
          |  |||  |||||| || |||  ||||||||| || |||||| |||||  ||||||| 
Db   1819 CAGGAGCAGGAGAGTTAGAGGAGGCAGGAGCAGGAGCAGGAGCAGGAGCAGTTAGAG 1878
```

```
RESULT 22
AC073609
ACCESSION   AC073609
LOCUS       AC073609                 211158 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Mus musculus clone RP23-518010, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
VERSION     AC073609.18  GI:16117946
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 211158)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,M., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,R., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Taminana,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211158)
```

```
Qy    718 TCCCTCAAAgAAAAgAGTATGAgAAATTGAAGGAAAgAGTCGGAGAGGTCGGAAGGCCACAGGGGAACTGGCT 777
          |||  | | ||| ||||| |  || ||| ||| | ||| ||| || |  ||| | | |  ||| || |
Db   1879 GAGCAGGAGCAGGAGAGTTAGAGGAGGCAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGTTA 1938

Qy    778 GACAGGTTGAAGAAGGAGATTTGGTGTCCTCTPAGGAGCAGGACAAGTTGAAGAGTTAGAGGAGCAGCACTGAG 837
          |||   || | ||  || ||||| || | || ||||||| || |||| || || ||| ||||||| || || 
Db   1939 GAGGAGCAGGAGCAGGAGGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGCAGGAG 1998

Qy    838 CTGGATCAGGCCAAGTTAGAACTGAGGTCAGCCCAGAAGGAGACTTACAAAGTGCTGACCAG 897
          ||||| |||| | ||||||   || |||  |||| |||| |||| ||    || |   ||| 
Db   1999 TTAGATGAGCAGGAGGAGCAGGAGGAGTTAGAGGAGGAGCAGGAGGAGTTAGAGGAGCAGGAG 2058

Qy    898 GAGATCACGAGCGACCCTAAGAAAAACAAAGTCTGATGATC 931
          |||| |||| ||||| |||| | ||| |||| | || ||| 
Db   2059 GAGTTAGAGGAGCAGGAGCAGGAGCAGGAGTTAGAGGAGC 2092
```

```
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (27-JUN-2000) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Oct 14, 2001 this sequence version replaced gi:12043813.
           --------------------------------------------
           Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu/
           Contact: hgsc-help@bcm.tmc.edu
           --------------------------------------------
           Project Information
           Center project name: HBFY
           Center clone name: RP23-518010
           --------------------------------------------
           Summary Statistics
           Sequencing vector: M13; L08821
           Chemistry: Dye-primer BodIpy; 12% of reads
           Chemistry: Dye-terminator Big Dye: 88% of reads
           Assembly program: Phrap; version 0.990329First call to
           findPhrapList
           Consensus quality: 206544 bases at least Q40
           Consensus quality: 209406 bases at least Q30
           Consensus quality: 211582 bases at least Q20
           Estimated insert size: 209430; sum-of-contigs estimation
           Quality coverage: 0x in Q20 bases; agarose-fp estimation
           Quality coverage: 6.6x in Q20 bases; sum-of-contigs estimation
           --------------------------------------------
           * NOTE: Estimated insert size may differ from sequence length
           * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 6 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *
           *       1: contig of 69093 bp in length
           * 69094: gap of unknown length
           * 69193: contig of 52993 bp in length
           * 122187: gap of unknown length
           * 122286: contig of 33305 bp in length
           * 155591: gap of unknown length
           * 155592: contig of 32877 bp in length
           * 188568: gap of unknown length
           * 188669: contig of 15512 bp in length
           * 204181: gap of unknown length
           * 204280: contig of 6878 bp in length.
           * 204281:
           * 211158:
FEATURES             Location/Qualifiers
     source          1..211158
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-518010"
BASE COUNT    67652 a     39628 c     38593 g     64781 t     504 others
ORIGIN
```

Query Match   2.68;   Score 51.4;   DB 2;   Length 211158;
Best Local Similarity   45.3%;   Pred. No. 0.0041;
Matches 187;   Conservative   0;   Mismatches 226;   Indels   0;   Gaps   0;

```
Qy         488 ACAAGGCAGAGATGCTGTGTTCCACCCTGAAAAAACAGATGAAGTTCCTGGAGCAGCCGGC 547
               ||||  ||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db          26 AGAAGGAGGAGGAAGAGGGGGAAGGAGGAGGAAGAGGAGGAAGGAAGAGGAGAGAAGG     85

Qy         548 AGGATGAGACCAAAAACAGAGCTCGGGAGGAGGCCCACCGACTCAAGTGCAAGATGAAAACCA 607
               ||||| |||      ||  ||||||||||||||||||||| ||| | |||| |  |||||||
Db          86 AGGAGAAGGAGAAGGAGGAGGAGGAGGAGGAAGGAGGAGGAGAGAAGG                145

Qy         608 TGGACCAAAATTGAGCTCCTCCAGAGCCAGCGTGTTCTCAGGGTGGAGGAGGATGATTCGAG 667
               ||||         |  |||||||||||||||||||||||||||||||||||||||||
Db         146 AGGAGGAGGAGGAGAAGGAGGAGGAGGAGGAGGAGAAGGAGGAGGAGAGGAGGAGGAGG     205
```

```
RESULT  23
LOCUS       CHPINVOL               1683 bp    DNA     linear   PRI 27-APR-1993
DEFINITION  Chimpanzee (P.paniscus) involucrin, complete cds.
ACCESSION   M26514
VERSION     M26514.1  GI:176809
KEYWORDS    epidermal protein; involucrin.
SOURCE      P.paniscus (strain CHP) keratinocyte DNA.
  ORGANISM  Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 1683)
  AUTHORS   Djian,P. and Green,H.
  TITLE     Vectorial expansion of the involucrin gene and the relatedness of
            the hominoids
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 8447-8451 (1989)
   MEDLINE   90046814
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            submitted by P.Djian, 26-JUL-1989.
FEATURES             Location/Qualifiers
     source          1..1683
                     /organism="Pan paniscus"
                     /db_xref="taxon:9597"
     CDS             1..1683
                     /note="involucrin"
                     /codon_start=1
                     /protein_id="AAA35422.1"
                     /db_xref="GI:176810"
                     /translation="MSQQHTLPVLLSPALSQELKTVPPPVNTQQEQMKQPTPLPPPC
                     QKMPVELPVEPVPSKQEEKHHTAVKGLPEQDECEQQQQEPQSELQQQHWEQHEEYQKAE
                     NPEQQLKQEKAQRDPQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELLEPQQE
                     GHLKHLEQRGQLELPEQQGQLKHLEQQKGQLHLEQQGQLKHLEQQKGQLELPEQQGQLKHL
                     EQEGQPKHLEBOEGQLEVPEQQVGQLKYLEQQGQLKHLEQQKGQLEQLEHLEHOEGQLPEQQVGQLK
                     QLEKEEGQPKHLEBEEGQLHLVQQEQGQLKHLVOQEEGQLKHLQQEGQLEQQGQVEH
                     LEQQVEHLEQLGOLKHLEQQGQLHLEQQEGQLKHLEQQGQLEQPEQQODGQLKHLEQQGQLKHLEQ
                     QEGQLKHLEKQEAQLELPEQVGQPKHLEQGQLHLEQQGQLEPPEQOODGQLKHLEQQGQLKDL
                     EQQGQLEQPVFAPAPGQVQDIQSALPTKGEVLLPLEHQOQKOEVQWPRKHK"
ORIGIN
BASE COUNT    496 a     419 c     597 g     171 t
```

Query Match   2.68;   Score 51.2;   DB 9;   Length 1683;
Best Local Similarity   49.3%;   Pred. No. 0.0036;
Matches 134;   Conservative   0;   Mismatches 138;   Indels   0;   Gaps   0;

```
Qy         523 CAGATGAAAGTTCCTGAGCAGCCGGCAGGATGAGACCAAAAACAAAGCTCGGGAGGAGGCCCAC 582
               |||| ||||||||||||||||||  ||||||||| ||||   |||||||||||||||||||||
Db        1114 CAGCTCGAGCAATCTGGTGCAGCCAGAGACCCAGCTGGAGCAGGAGGGGGCAGGTGGAG    1173

Qy         583 CGACTCAAGTGCAAGATGAAAACCATGGAGCAAAATTGAGCTCCTCCAGAGCCAGCGCGT 642
               | ||||||  |||| |||| |||     ||||||| |||||||||||||||||||||||
Db        1174 CACCTGGAGCAGCAGGTGGAGCACCTGGACCAGCAGTTGGGCAGCACCTAGAGGAG    1233

Qy         643 TCTGAGGTGGAGGAGGAGTGATTCGAGACAATGGGTGGTGGGACAGTGACCGGTGGAGCAGCTG 702
               ||||| |  ||||||||||||  ||||      |||  ||||||||||| ||| ||||| | |
Db               [alignment]
```

```
Db  1234  CAGGAGGGGCAACTGACCATCTGGAGCAGCAGCAGGACAGTTGGGGGGTCCCAGAGCAG 1293
          |||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   703  GCTGTGTACTGCGTGTCCCTCAAGAAAGGAGTATCAGAATCTGAAGGAAGCTCGGAAGCC 762

Db  1294  GTGGGGCAACCAAAAGAACTGGAGCAGGAGGAGGGAGAAGCAGCTGGAGCTCCCAGAGCAGCAGCAA 1353
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   763  ACAGGGGAAACTGGCTGACGAGGTTGAAGAAGGA 794

Db  1354  GAGGGGCAGCTGAAGCACCTGGAGAAAGCAGGA 1385
          ||||  ||||||||||||||||||||||||| |||
Qy  1354
```

RESULT 24
LOCUS   U93872
LOCUS   U93872/c
DEFINITION   U93872   133661 bp   DNA   linear   VRL 09-JUL-2001
             Kaposi's sarcoma-associated herpesvirus glycoprotein M, DNA
             replication protein, glycoprotein, DNA replication protein, FLICE
             inhibitory protein and v-cyclin genes, complete cds, and tegument
             protein gene, partial cds.
ACCESSION   U93872
VERSION   U93872.2 GI:14627174
KEYWORDS   .
SOURCE   Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.
ORGANISM   Human herpesvirus 8
             Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
             Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1 (bases 29032 to 30108; 117733 to 118431)
AUTHORS   Zhong,W., Wang,H., Herndier,B. and Ganem,D.
TITLE   Restricted expression of Kaposi sarcoma-associated herpesvirus
             (human herpesvirus 8) genes in Kaposi sarcoma
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)
MEDLINE   96270595
REFERENCE   2 (bases 17242 to 17856)
AUTHORS   Neipel,F., Albrecht,J.C., Ensser,A., Huang,Y.Q., Li,J.J.,
             Friedman-Kien,A.E. and Fleckenstein,B.
TITLE   Human herpesvirus 8 encodes a homolog of interleukin-6
JOURNAL   J. Virol. 71 (1), 839-842 (1997)
MEDLINE   97138401
REFERENCE   3 (bases 123309 to 124082)
AUTHORS   Li,M., Lee,H., Yoon,D.W., Albrecht,J.C., Fleckenstein,B., Neipel,F.
             and Jung,J.U.
TITLE   Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin
JOURNAL   Nature 386 (6624), 517-521 (1997)
MEDLINE   97242415
REFERENCE   4 (bases 122660 to 123226)
AUTHORS   Thome,M., Schneider,P., Hofmann,K., Fickenscher,H., Meinl,E.,
             Neipel,F., Mattmann,C., Burns,K., Bodmer,J.L., Schroter,M.,
             Scaffidi,C., Krammer,P.H., Peter,M.E. and Tschopp,J.
TITLE   Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced
             by death receptors
JOURNAL   Nature 386 (6624), 517-521 (1997)
MEDLINE   97296220
REFERENCE   5 (bases 1 to 133661)
AUTHORS   Neipel,F., Albrecht,J.C. and Fleckenstein,B.
TITLE   Cell-homologous genes in the Kaposi's sarcoma-associated
             rhadinovirus human herpesvirus 8: determinants of its
             pathogenicity?
JOURNAL   J. Virol. 71 (6), 4187-4192 (1997)
MEDLINE   97184528
REFERENCE   6 (bases 1 to 133661)
AUTHORS   Neipel,F., Albrecht,J.C., Ensser,A., Huang,Y.-Q., Li,J.J.,
             Friedman-Kien,A.E. and Fleckenstein,B.
TITLE   The genome of human herpesvirus 8 cloned from Kaposi's sarcoma
JOURNAL   Unpublished
REFERENCE   7 (bases 1 to 133661)
AUTHORS   Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
             Friedman-Kien,A.E. and Fleckenstein,B.
TITLE   Direct Submission
JOURNAL   Submitted (17-MAR-1997) Virology, University of Erlangen,
             Schlossgarten 4, Institut for Klinische und Molekulare, Erlangen
             91054, Germany
REFERENCE   8 (bases 1 to 133661)
AUTHORS   Neipel,F., Albrecht,J.-C., Ensser,A., Huang,Y.-Q., Li,J.J.,
             Friedman-Kien,A.E. and Fleckenstein,B.
```

TITLE   Direct Submission
JOURNAL   Submitted (09-JUL-2001) Virology, University of Erlangen,
             Schlossgarten 4, Institut for Klinische und Molekulare, Erlangen
             91054, Germany
REMARK   Sequence update by submitter
COMMENT   On Jul 9, 2001 this sequence version replaced gi:2246466.
FEATURES   Location/Qualifiers
     source   1..133661
                     /organism="Human herpesvirus 8"
                     /db_xref="taxon:37296"
                     /note="sequence derived from overlapping Lambda phage
                     clones isolated from two libraries of Sau3A partially
                     digested DNA from Kaposi's sarcoma biopsies"
     CDS   105..959
                     /note="ORF K1"
                     /codon_start=1
                     /protein_id="AAB62619.1"
                     /db_xref="GI:2246494"
                     /translation="MFLYVVCSLAVCFRGLLSLSLQSSPNLCPGVISPYTLTCPSNT
                     SLPTSWYCNDTRLRLTQQTFTVDTLICNFSCVQQSGHRHSLNITWYPQVLGTLCGQ
                     PSNVTCGQHVTLYCSTSCGNNVTAVWHLDNGRNEIVSQTKYYNFTLMNQTEGCYACSNG
                     LSSRLSNRLCFSARCANLTPETHTVSVSSTTGTRFETFSNRRGTTLDLVLVMKEKSTNL
                     HIQVHFLVFMTLVALIGIMCGLIGTIIFAHCQKQRDSNKTVPQQLQDYYSLHDICTED
                     YTQPVDWY"
     CDS   complement(1127..2779)
                     /note="ORF 04, complement control protein homolog"
                     /codon_start=1
                     /protein_id="AAB62602.1"
                     /db_xref="GI:2246477"
                     /translation="MAFLRQTLWTLWTFTMVIGQDNEKRCKSQKTLIGYRLKMSRDGDIA
                     QDALKYGANISVCNEGYFLVGREYVRVCMIGASGQMAWSSSSPPFCEREKCHPREKIEN
                     GDFKPDKDYYEYNDAVHFECNEGYTLVGPHSIACAVNNTWISMNPTCELAGCKFPSVT
                     HGYPIQGRSLTYKIHKQSVTFACNDGYKLLPSPITCNYTEWDDPLPKVLEDIDDPNN
                     SNPGRLHPPNPNEKRGNVEQRSNVTEPPFKPEDTHTAATCDTNCEQPRILPTSEGEN
                     ETTTSNTITKQLEDEKTTSOPNTHITSALTSMKAKGNFTNKTNNSTDLHIASTPTSQD
                     DATPSIPSVQTPNYNTNNAPRTLTTLSLHIEEGPSNSTTSEKATASTLSHNSHKNDTGGI
                     YTTLNKTTQLPSTNKPTNSQAKSSTKPRVETHNKTTSNPAISITDSADVPQRPREPTL
                     PPIFRPPASKNRYLEKQLVIGLLFAVALICGLITLFHYLFFR"
     CDS   3194..6592
                     /note="ORF 06, major ssDNA binding protein homolog"
                     /codon_start=1
                     /protein_id="AAB62603.1"
                     /db_xref="GI:2246478"
                     /translation="MALKGPQTLEENIGSAAPTGPCGYLIYAYLTHNFPIGEASLIGNG
                     YPEAKVFSLPLLIGITLVESDFPLNVKAVHKKILDATTTASVKLTSYHREAIVFHNTHLFQ
                     PIFQGKGLEKLCRSSRELEGFSPTVEQVNAVHKHILPSSPEACPLQCANEIFMAVIVTEGF
                     KERLYGGKLVPVPSQTTPVHIGEHQAFKIPLYDEDLFGPSRAQELCRFEYNPDISRYLH
                     DSIFTGIAQALRVKDVSTVIQAEBRQFVHDQYKIPKLVQAKDFPQCASRGRQTDCSTLMV
                     IDSIVAELGMSVGLSFIEGPSGCEVLNVDTWFIFENCETPDARLRALEVWHRQALH
                     IGAQLFAANSVLLTRVAKLPQKNQRGDANMYNSFYLQHGLGYLSEATVKENGASAFK
                     GVPVSALDGSSYTLQHLAYASSSSPHLLARMCVYLQFIPHHKHYNSOSYNVAPYATT
                     APSQMCDLCQGQCPAVCINTLFYRMKDRFPPVLSNVRKNRDPYVITGTAGTYNDLEIGN
                     FATREREEEGNPVEDAPKYTWOLCQNITEKLASMGISEGGDALRTLLIVDIPSFKV
                     FKGIDSTVEAELLKFINCMIKNIVNVNFRNIKSVHHILQPACNYWQACPVFLIYYK
                     SLIVIODICILTSCMMYEQDNPAVGIVPSEWLKMHFQTMWTNFKGACFDKGATGGEL
                     KIVHQSMFCDLFDTDAAIGGMFAPARMQVRIARAMLMVPKTIKIKNRIIFSNTGAES
                     IQAGFMKPASQRDSYIVGGPYMKFLNALHKTLPSTKTSALYLWHKICGQTQPYATT
                     VSGEHLTELCNYVKASSQAFEEINVLDLVPDTITYSAKILNSSILRACGQTQPYATT
                     LSCLSPVTQLVPAEEYPHVLGPVCLSSDDEYRAKVAGRSVTIVQSTLKQAVSTNGRLR
                     PITYVPLVVNKYGYSGVDRNLRPESVPFRKNNVSSMLRKR
                     HVIMPLVDRLVKRIVGINSGEFEAEAVARSVQNVLEDRDNPNLPKTVVLELVKHLGS
                     SCASLTEEDVIYLGPYAVLGDEVLSLLSTVGQAGVPWTAEGVASVIQOIIDDCELQF
                     VGPEEPCLLQGQSVVEELFPSPGVPSLTVGKKRIASLLSDLDL"
     CDS   6609..8696
                     /note="ORF 7, processing and transport protein homolog
                     (ICP18.5)"
                     /codon_start=1
                     /protein_id="AAB62620.1"
                     /db_xref="GI:2246495"
                     /translation="MAKELAAVIADVSALAMDLCLLSTADPATLDTKSLALTGKFEQS
                     LHGTLLPLLRRQNAHECSCLSLELEHLLENVADALATLGVCTSRKLSPEEHFSLLHLD
                     TTCNKHRSVRFNFYGNWALELKLSLINDVEIFERKLRLSSVFYCICGSGSALEGLGEVLRF"

```
RESULT 25
AL589701/c
LOCUS       AL589701       219200 bp    DNA    linear    ROD 30-JAN-2002
DEFINITION  Mouse DNA sequence from clone RP23-202F3 on chromosome 13.
ACCESSION   AL589701
VERSION     AL589701.9  GI:18476660
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 219200)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:17384104.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source database entries:
            SWISSPROT: Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202F3 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-202F3 The true
            left end of clone RP23-171015 is at 179956 in this sequence. The
            true right end of clone RP23-153B6 is at 117857 in this sequence.
FEATURES             Location/Qualifiers
     source          1..219200
```

Query/DB alignment (Db 125729 ... Qy 881):

```
Db 125729 AGCAGCAGGAGGAGCAGGAGCAGGAGGCAGGAGGCAGCAGGAGCAGGAGGCAGGAGGCAGGAGGAGCAGGAGGCAGG 125670
Db 125669 AGCAGGAGGAGGAGCAGGAGGCAGGAGGCAGCAGG ... 125610
...
Qy 881 TACAAAGTGCTGACCAGGAGATCACGAGCCTAAGAAAAGAAGTCTGATGATC 931
Db 125369 TAGAGGAGCAGGAGCAGGAGGAGGAGTTAGAGGAGC 125319
```

CDS features:

```
CDS
     /note="ORF 08, glycoprotein B (gB) homolog"
     /codon_start=1
     /db_xref="GI:2246592.1"
     /protein_id="AAB62592.1"
     /translation="MTPRSRLATLGTVILLVVCFCAGAAHSRGDIFQTSSSPPGSSS
     KAPTKPGEEASGEKSVDEYQFRVCSASITGELERFENLEQTCPDIKDKYHQEGILLVYK
     KNIVPHIFKVRYRKIATSVTVIVRGLTESAITNKYELEPRPVPLYEISHMDSTYQCFSS
     MKVNVNGVENTFDRDDVNTTVLELGLTDNIORYFESOPVIVAEPGNFGGPIVRVRT
     TVNCEIVDMIARSAEPYNYFVTSLGDTVEVSPFCYNESSCSTTPSNKNGLSVQVVLNH
     TVVTYSDRGTSPTPQNRIFVETGAYTLSWASESKTTAVCPLALMKTFPRSIOTTHEDS
     FHFVANAEITATFAPLTPVVANFIDTIYSCLTSDINTTLNASKALASTHVPNGTVQYFH
     GSTDNLSYTQLQFAYDKLRDGINQVLEELSRANCREQVRDNLMWYELSKINPTSVMTA
     IYGRPVSAKFVGDIAISVTECINVDQSSVNIHKSLRTNVSDVTARPLVTFKFLNSSNL
     FTGOLGARNEIILLINNQVETCKDTCEHVFITRNETLVKDYAYLRTINTTDISTLNTF
     IALNLSFIQNIDFKAIELYSSAEKRLASSVFDLETMFREYNYTTHRLAGLREDLDNTI
     DMNKERFVRDLSEIVADLGGGIGKTVNVNASSVVTLCGSLVTGFINFIHPLGMLII
     IVIAILLIIFMLSRRTNTIAQAPVKMIYPDVDRRAPPSGGAPTREEIKNILLGMHQLQ
     QEERQKADDLKKSTPSVFQRTANGLRORLRGYKPLTOSLDISPETGE"

CDS
     /note="ORF 09, DNA polymerase homolog"
     /codon_start=1
     /db_xref="GI:2246468.1"
     /protein_id="AAB62593.1"
     /translation="MDFFNPFIDPTRGGPRNTVRVQRPTPSQSPIVPSETRVCRLIPACF
     QTPGRPGVWAVADVTFPPFYFQGPKRGEVFAGETGSIWKTRRGGARNAPMSHLIFHVYD
     IVETTYTADRCEDVPFSFOTDIIPSGTVLKLLGRTLDGASVCNVNFRORCYFYTLAPQ
     GVNITHVLOQALQAGFGRASCGFSTEPVRKKILRAYDTOQYAVQKITLSSSPMMRTLS
     DRITTCGGEVFESNVDAIRRFVLDHGESTFGWECSNPARTOARDSWTELEFDCSWE
     DLKFIPERWEPPYTILSEFDIECMGEKGFPNATODEDMIIQISCVLHYVGNDKPYTRM
     LLGLGTCDPLPGVEVEFEEPSEVDMLAAFLSMLRDYNVEFITGYNIANANFDLPYIIARAI
     QVDFKLQDFTKIKTGSVEVEVHQPRGGSDGGNPEMRSQSRKVIGGLFKLQGVEPEIDV
     SLSDYKLDIVAKQCLGROKDDISYKDIPPLFKSGPDGRAKVGNVCVIDSVLVMDLLR
     FQTHVEISEIAKLAKIPTRRVLTDGQOIRVFSCLLEAAATEGYILPVRGDAVSGYQG
     ATVISPSGFYDDPVLVDFALYPSIIQAHNICYIKLIRAYDIQOYAVQKITLSSSPMMRTLS
     VLSGGPVHFVKHKRESLLAKLLTVWLAKRKEIRKTLASCTDPALKTTILDKQOLAIKV
     TCNAVGFTGVASGILPCLNIAETVTLQGRKMLERSOAFVEAISPERLAGLLRRPIDV
     SPDARFKVIYGDTDSLFICCMGENMDSVSDFAEELASITITNTLFRSPIKLEAEKIFKC
     LLLLTKKRYVGVLSDDKVLMKGVDLIRKTACRFVOEKSSOVLDLIIREPSVKAAAKLI
     SOOATDWVYREGLPEGFVKIIQVLNASHRELCERSVPVDKLTFTTELSRPLADYKTON
     LPHLTVTOKLQOAROEELPOIHDRIPYVVDAPGSLRSELEAEHPEYVKQHGLRVAVDLY
     FDKLVHAVANIIQCLFQNNTSATVAMLVNFLDIPVTFPYP"

CDS
     /note="ORF 10, herpesvirus saimiri ORF 10 homolog,
     conserved in other gamma-herpesviruses"
     /codon_start=1
     /db_xref="GI:2246469.1"
     /protein_id="AAB62594.1"
     /translation="MQTEATFILGDWEITVSNCRFTCSSLTCGPLYRSSGDYTRLRIP
     FSLDRLIRDHAIFGLVPNIEDLLTHGSCVAVADANARGNARIVAPGVINFSEPI
     GIWVRGPPPQTRKEAIKPCIFYEVSPLPREMTTVVFKGGDLPPGAEEPETLHSAEAPL
     PSRETLVTGQLRSTSPRIYTGYTGYFHSPVVPLSFDLLTFESIGCDNVEGDPEOLTPKYLT
     FQTGERLCKAVRAVYNTHSTACKRARVRVRYRVRPFSAVRVGAQASPLITPLGCARVFA
     VPDCEKIIPPOETTLRIQLLFEOHGANAGDCAFVIGLARETKEVSFPAVLLPGKH
     EHLIVFNPQTHPLTIORDTIVGVAMACYIHPGRAASQAPYSFYDCKEESWHVGLFQIK
     RGPGGVCTPPCHVAIRADRHEEPMQS"
```

Left column (query match statistics):

```
Qy   521 AACAGATGAAGTTCCTGAGCACGCACGGCAGGATGAGACCAAAACAAGCTCGGGAGGAGGCCC 580

Query Match   45.3%;   Score 51;   DB 14;   Length 133661;
Best Local Similarity   45.3%;   Pred. No. 0.0052;
Matches 186;   Conservative   0;   Mismatches 225;   Indels   0;   Gaps   0;

2.68;
```

misc_feature
/organism="Mus musculus"
/db_xref="taxon:10090"
/note="Single clone region"
/chromosome="13"
/clone="RP23-202F3"
/clone_lib="RPCI-23"
3942..4127
misc_feature
2978..3099
/note="Single clone region. Sequence from clone PCR only."
/note="Single clone region. Sequence from reads from a
short insert library derived from a clone PCR. Restriction
digest data confirm the assembly."

BASE COUNT   63635 a   46953 c   46966 g   61646 t
ORIGIN

Query Match   2.6%;   Score 50.8; DB 10; Length 219200;
Best Local Similarity 44.6%; Pred. No. 0.0061;
Matches 199; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

Qy   476 AGAACGCCTTAAACAAGGCAGAGAGATGCTCGTGTTCCACCCTGAAAAAACAGATGAAGTTCC 535
          |||| ||| || | | ||||| |||| | | || |||||||||||| |||||||||
Db 61398 AGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGATTTGGTGTTCCTCTAGGAGACTCTCAA 61339

Qy   536 TGGAGCAGCGCGGCCAGGATGAGACCAAACAAGCTCGGGAGGAGGCCCACCGACTCAAGTGCA 595
          ||| || || | || || |||||| ||||| | |||||| || ||| |||||
Db 61338 AGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAAGG 61279

Qy   596 AGATGAAAAACCATGGAGCAAATTGAGCTCCTTACTCCAGAGCCACCGTTCTGAGGTGGAGG 655
          || | |||| ||||||| ||| | ||| | ||||| |||||||| || || |||||||
Db 61278 AGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAAGG 61219

Qy   656 AGATGAATTCCAGACACATGGGTGTGGGGACAGTTCAGCCGTTGAGCTGGCTGTGTGTACTGCG 715
          || | || |||||||||||| ||| ||| ||| |||| ||||| ||||| || || ||||
Db 61218 AGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAAGA 61159

Qy   716 TGTCCCTCAAGAAAAGAGTATGCAGAATCTGAAGGAAGCTCGGAAGGCCACAGGGAACTGG 775
          | | | |||||||| | ||| |||||||||||||| |||||| |||||| ||||
Db 61158 AGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGG 61099

Qy   776 CTGACAGGGTTGAAGAAGGATTTGGTGTCCTCTAGGAGACTCTCAACACTG 835
          | |||||| |||| ||||| |||||| || ||||||||||| ||| ||
Db 61098 AGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGG 61039

Qy   836 AGCTGGGATCAGGCCCAAGTTAGAACTGAGGTGTCAGCCCAGAAGGAAGTTGCTGACC 895
          ||| ||| ||||| | |||||||| ||||| ||||| |||| |||||||| || |||
Db 61038 AGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAGGAAGGAAGAGGAAGG 60979

Qy   896 AGGAGGATCACGAGCCTTAAGAAAGAAG 921
          | |||| | | | |||| ||||||
Db 60978 AGGAGGAAGGAGGAAGGAGGAAGGAAGAGGAAGG 60953

RESULT 26
LOCUS   AF360120   3012 bp   DNA   linear   VRL 13-AUG-2001
DEFINITION Human herpesvirus 8 ORF73 gene, complete cds.
ACCESSION   AF360120
VERSION   AF360120.1 GI:13936995
KEYWORDS   .
SOURCE   Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.
  ORGANISM Human herpesvirus 8.
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1 (bases 1 to 3012)
  AUTHORS   Garber,A.C., Shu,M.A., Hu,J. and Renne,R.
  TITLE   Dna binding and modulation of gene expression by the
            latency-associated nuclear antigen of Kaposi's sarcoma-associated
            herpesvirus
  JOURNAL   J. Virol. 75 (17), 7882-7892 (2001)
  MEDLINE   21376412
   PUBMED   11483733
REFERENCE   2 (bases 1 to 3012)
  AUTHORS   Garber,A.C., Gonzalez,D.Y. and Renne,R.

  TITLE   Direct Submission
  JOURNAL   Submitted (13-MAR-2001) Molecular Biology and Microbiology, Case
            Western Reserve University, 2109 Adelbert Rd., Cleveland, OH 44106,
            USA
FEATURES   Location/Qualifiers
     source   1..3012
                     /organism="Human herpesvirus 8"
                     /db_xref="taxon:37296"
                     /note="from KS long tumor"
     CDS   1..3012
                     /codon_start=1
                     /product="ORF73"
                     /protein_id="AAK50002.1"
                     /db_xref="GI:13936996"
                     /translation="MAPPGMRLRSGRSTGAPLTRGSCRKRNRSPERCHLGDDLHLQPR
                     RKHVADSVDGRECGPHTLLPIGSPTVFTSGLPAEVSSPPLPVAPIPSPAPATLPPPA
                     LLPVTTSSSPIPPSHPVSPGTTDHSPSPALPPPQSPRSSPDDNTLSIQSSQQQQEDPQ
                     MRPPPSQQTTPPHSPTTPPEPPSKSSPDSLAPSTLRSLRKRRLSSPQGPSTLNPICQ
                     SPPVSPPRCDPFANRSVYIPWNATESPIYVGSSSSDDGDTPPTPSTISIGSSSVPSEGSW
                     GDDTAMLVLLAEIAREASKNEKECSENNQAGENGDNEISKESOVDKDNDNDKDDEEE
                     QETEDEDEEDNEDEEEDKKEDEDEDGDNKTLSIQSSQQQQEDPQMQQEPEQQQEPQQQ
                     DEEDDDEDNEDEEEDKKEDEDEDGDNKTLSIQSSQQQQEDPQMQQEPEQQQEPQQQDE
                     QQEPLQEPQQOEPQQQEPQQQEPQQQEPQQQEPQQQEPQQQOEDPQMQQEPEQQQEPQ
                     QQEDQEPQQEPQQEPQQEPQQEDQEPQQEDQEPQQEDQEPQQEDQEPQQEDQEPQQE
                     EQQEPLEEQQEPLEEQQEPLEEQQEPLEEQQEDPQMQQEPEQQQEPQQQEPQQQDEQ
                     ELEEQQEPLEEQQEPLEEQQEPLEEQQEPLEEQQEPLEEQQEPLEEQQEPLEEQQEQ
                     EQELEEVEEQOEQELEEVEEQQOGVVVNQDTSEDDVEVEEPIILHGSSEDEMVDYFVVST
                     HEQIASSPPGDNTPDDDPQGPSREYRYVLRTSPPHRPQVRMRRVPVTHPKKPHPRYQ
                     QPPVPYRQIDDCPAKARPQHIFYRRFLGKDGRDPKCQWKFAVIFWGNDPYGLKKLSQ
                     AFQFGGVKAGPVSCLPHPGPDGSPITYCVVYVYCONKDTSKKVQMARLAWEASHPLAGN
                     LQSSIVKFKKFPLPIIQPGENQGPGDSPQEMT"
BASE COUNT   908 a   714 c   966 g   424 t
ORIGIN

Query Match   2.6%;   Score 50.4; DB 14; Length 3012;
Best Local Similarity 45.3%; Pred. No. 0.0064;
Matches 183; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

Qy   528 GAAGTTCCTGGAGCACGCGGCAGGATGAGACCAAACAAGCTCGGGAGGAGGCCCACCGACT 587
          |||||| || || |||| ||| ||| || | || ||| || ||||| ||| ||||
Db   1572 GCAGCAGCAGGATGACCAGCAGCCAGAGCCCCAACAGGAGCCAGAGCAGCAGGAGCAGGAGCA 1631

Qy   588 CAAGTCGAAGATGAAAAACCATGGAGCAAAATTGAGCTCCTTACTCCAGAGCCACCGTTCTGA 647
          ||| | ||| ||||||| || |||| ||||||||| | |||| || ||||||||| |||
Db   1632 GCAGGAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGA 1691

Qy   648 GGTGGAGGAGATGATTCGAGACATGGGTGTGGGGACAGTCACGGTGGAGCAGCTGCCTGT 707
          || |||||| | | ||||| | ||||| || |||| |||| ||| ||||| |||| ||||
Db   1692 GTTAGAGGAGCAGGAGCAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCA 1751

Qy   708 GTACTGCGGTGTCCCTCAAGAAAAGAGTATGCAGAATCTGAAGGAAGCTCGGAAGGCCACAGG 767
          || | | ||| |||| ||||||| |||| | |||||||||||||||| |||||| |||||
Db   1752 GGAGTTAGAGGAGCAGGAGCAGGAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCA 1811

Qy   768 GGAACTGGCTGACAGGGTTGAAGAAGGATTTGGTGTCCTCTAGGAGACTCTCAACAGTTGAAGACTCT 827
          |||| || || ||||| |||| ||||| |||||| || |||||||||||| || | |||||| |||
Db   1812 GCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCA 1871

Qy   828 CAACACTGAGTCTGGGATCAGGCCCAAGTTAGAACTGAGGTCACCCCAAGAGGACTTCACAAAG 887
          || || |||| ||| | ||| | |||||||| ||||| | ||| ||| |||||| | |||
Db   1872 GGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGA 1931

Qy   888 TGCTGACCAGGAGGATCACGACGACGCCTAAGAAAGAAGTCTGATGATC 931
          |||| |||||| || ||| | || || ||| |||||||| || | |||
Db   1932 GCAGGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGTTAGAGGAGC 1975

RESULT 27
LOCUS   AC007461   180385 bp   DNA   linear   PRI 04-JUN-1999
AC007461/c
DEFINITION Homo sapiens chromosome 17, clone 84_E_24, complete sequence.

```
ACCESSION   AC007461
VERSION     AC007461.8  GI:4996920
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180385)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 180385)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 4, 1999 this sequence version replaced gi:4966400.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Only 180385 base pairs from the middle of this clone are being
            submitted. The remainder overlaps either accession number AC005152
            (WICGR project L335) or accession number AC005243 (WICGR project
            L343).
FEATURES             Location/Qualifiers
     source          1..180385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="84_E_24"
                     /clone_lib="Alan Buckler -- per comm"
     repeat_region   446..753
                     /rpt_family="AluSc"
     repeat_region   2375..2419
                     /rpt_family="(TTTTG)n"
     repeat_region   complement(2985..3103)
                     /rpt_family="L2"
     repeat_region   complement(3146..3352)
```
```
     repeat_region   complement(3531..3639)
                     /rpt_family="L2"
     repeat_region   complement(3798..4394)
                     /rpt_family="L2"
     repeat_region   5531..5631
                     /rpt_family="MIR"
     repeat_region   5648..5674
                     /rpt_family="MER81"
     repeat_region   5812..6020
                     /rpt_family="(CAAAAA)n"
     repeat_region   6180..6224
                     /rpt_family="LIMC4"
     repeat_region   6303..6611
                     /rpt_family="AT_rich"
     repeat_region   7378..7618
                     /rpt_family="AluSp"
     repeat_region   9593..9623
                     /rpt_family="MIR"
     repeat_region   complement(10666..10923)
                     /rpt_family="(GGGA)n"
     repeat_region   10958..11088
                     /rpt_family="AluSq"
     repeat_region   11752..11914
                     /rpt_family="MIR"
     repeat_region   complement(11915..12219)
                     /rpt_family="L2"
     repeat_region   12220..12347
                     /rpt_family="AluSx"
     repeat_region   13110..13182
                     /rpt_family="L2"
     repeat_region   13526..13564
                     /rpt_family="GA-rich"
     repeat_region   13575..13660
                     /rpt_family="(TCC)n"
     complement(14380..14658)
                     /rpt_family="(TA)n"
     repeat_region   14742..14830
                     /rpt_family="AluJo"
     repeat_region   14860..14967
                     /rpt_family="(TA)n"
     complement(14977..15122)
                     /rpt_family="(TA)n"
     complement(15292..15385)
                     /rpt_family="L2"
     repeat_region   16672..16781
                     /rpt_family="L2"
     repeat_region   16802..16846
                     /rpt_family="L2"
     repeat_region   17225..17245
                     /rpt_family="(TGAA)n"
     repeat_region   18411..18720
                     /rpt_family="(A)n"
     complement(20715..20788)
                     /rpt_family="AluJb"
     complement(20822..20916)
                     /rpt_family="MIr1J2"
     repeat_region   20956..21137
                     /rpt_family="MLT1J2"
     repeat_region   21830..21872
                     /rpt_family="MIR"
     repeat_region   21957..22257
                     /rpt_family="AT_rich"
     complement(22438..22538)
                     /rpt_family="AluSx"
     complement(23375..23667)
                     /rpt_family="MLT1J"
     repeat_region   24553..24858
                     /rpt_family="AluY"
     complement(25021..25234)
                     /rpt_family="AluY"
     repeat_region   26957..27261
                     /rpt_family="MIR"
                     /rpt_family="AluY"
```

```
repeat_region        complement(27524..27669)
                     /rpt_family="MLT1B"
repeat_region        complement(27670..27964)
                     /rpt_family="AluSp"
repeat_region        complement(27965..28210)
                     /rpt_family="MLT1B"
repeat_region        complement(28228..28522)
                     /rpt_family="AluSg"
repeat_region        complement(28592..28700)
                     /rpt_family="MIR"
repeat_region        complement(29353..29537)
                     /rpt_family="MIR"
repeat_region        29802..30095
                     /rpt_family="MIR"
repeat_region        complement(30099..30219)
                     /rpt_family="AluSg"
repeat_region        30227..30267
                     /rpt_family="MIR"
repeat_region        30276..30321
                     /rpt_family="(TG)n"
repeat_region        complement(30322..30611)
                     /rpt_family="(CA)n"
repeat_region        complement(30956..31151)
                     /rpt_family="AluJo"
repeat_region        complement(31533..31633)
                     /rpt_family="MER3"
repeat_region        31810..31858
                     /rpt_family="L2"
repeat_region        32028..32107
                     /rpt_family="L2"
repeat_region        complement(32919..33182)
                     /rpt_family="L1PA4"
repeat_region        34169..34314
                     /rpt_family="MIR"
repeat_region        34774..35007
                     /rpt_family="CT-rich"
repeat_region        35589..35621
                     /rpt_family="MIR"
repeat_region        35763..36142
                     /rpt_family="AT_rich"
repeat_region        36143..36452
                     /rpt_family="MER41A"
repeat_region        36453..36636
                     /rpt_family="AluSg"
repeat_region        38011..38071
                     /rpt_family="MER41A"
repeat_region        38083..38262
                     /rpt_family="CT-rich"
repeat_region        38235..38319
                     /rpt_family="(TTC)n"
repeat_region        complement(38499..38820)
                     /rpt_family="(TTC)n"
```

Query Match   2.5%;   Score 50.2; DB 9; Length 180385;
Best Local Similarity   44.6%;   Pred. No. 0.009;
Matches 196; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

```
Qy     491 AGGCAGAGAATGCTGTGTTCCACCCTGAAAAAAACAGATGAAAGTTCCTGGAGCAGGGCAG 550
           ||| | |||| |||   ||||||||      |||| | |||||||   || ||||||| ||
Db   38515 AGGAGAGAAGGAGGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGGGGCAG 38456

Qy     551 ATGAGACCAAACAAAGCTCCGGGAGGAGGCCCACCGCACTCAAGTCGAAGATGAAAACCATG 610
           | |||      || ||   ||||| ||  | || |  |  ||  ||||| |||| |  | |
Db   38455 AGGAGGGAGAAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 38396

Qy     611 AGCAAATTGAGCTCCTACTCCAGACACCAGCGTTCTGAGGTGGAGGAGATGATTCCAGACA 670
           |||      || ||   || ||  |    |    ||   ||||||||| |  |||| | ||
Db   38395 AGAAGAAGAAGGAAGGAGGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 38336

Qy     671 TGGGTGTGGGACACTCACCGGGTGCGAGCACCTGGCTGTCTGTTGTTACTGCCGTTCCTCAAGAAAG 730
           ||                |  ||| |      ||  | | || | |  | ||| | | |||  ||
Db   38335 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAGGAGG 38276
```

RESULT 29
LOCUS   AF148805   28559 bp   DNA   linear   VRL 02-AUG-1999
DEFINITION   Kaposi's sarcoma-associated herpesvirus ORF 68 gene, partial cds;
             and ORF 69, kaposin, v-FLIP, v-cyclin, latent nuclear antigen, ORF
             K14, v-GPCR, putative phosphoribosylformylglycinamidine synthase,
             and LAMP (LAMP) genes, complete cds.

ACCESSION   AF148805
VERSION   AF148805.1 GI:5669888
KEYWORDS   .
SOURCE   Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.
  ORGANISM   Human herpesvirus 8
             Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
             Gammaherpesvirinae; Rhadinovirus.
REFERENCE   1 (bases 1 to 28559)
  AUTHORS   Glenn,M., Rainbow,L., Aurad,F., Davison,A. and Schulz,T.F.
  TITLE   Identification of a spliced gene from Kaposi's sarcoma-associated
             herpesvirus encoding a protein with similarities to latent membrane
             proteins 1 and 2A of Epstein-Barr virus
  JOURNAL   J. Virol. 73 (8), 6953-6963 (1999)
   MEDLINE   99329221
REFERENCE   2 (bases 1 to 28559)
  AUTHORS   Glenn,M.A., Rainbow,L., Aurade,F., Davison,A.J. and Schulz,T.F.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-APR-1999) MRC Virology Unit, Church Street, Glasgow
             G11 5JR, UK
FEATURES   Location/Qualifiers
     source   1..28559
                     /organism="Human herpesvirus 8"
                     /isolate="GK18"
                     /db_xref="taxon:37296"
                     /note="right end of the unique region plus 8 complete
                     copies and 1 partial copy of the terminal repeat (TR)"
     gene   <1..2105
                     /gene="ORF 68"
     CDS   <1..1102
                     /gene="ORF 68"
                     /function="involved in DNA packaging"
                     /note="similar to HSV-1 UL32, HCMV UL52 and EBV BFLF1"
                     /codon_start=2
                     /evidence=not_experimental
                     /protein_id="AAD46496.1"
                     /db_xref="GI:5669889"
                     /translation="IMHLHFPEEEATYLGLFGPGPRLMGIDLQLHFFVQKCFKTTAAEKI
                     LGISNLQFIKSEFIRGMLTGTITCNFCFKTSWPRDKEBATGPPCCQITDTTTAPAS
                     GIPELARATFCGASRPTKFSLLPALIDIWSTSSELLDEPRRLIASDMSELKSVASH
                     DPFFSPPLQADTSQGPCLMHPTLGLRYRKNGTASVCLLCECLAAHPEAPKALQTLQCEV
                     MGHIENNVKLVDRIAFVLDNPFAMPYVSDPLLRELIRGCTPQEIHKHLPCDPLCALNA
                     KVVSEDVLFRLPREQEYKKLRASAAAAGQLLDANTLFDCEVVQTLVFLFRGLQNARVGK
                     TTSLDIIRELTAQLKRHRLDLAHPSQTSHLYA"
     gene   1135..2105
                     /gene="ORF 69"
     CDS   1135..2043
                     /gene="ORF 69"
                     /note="similar to HSV-1 UL31, HCMV UL53 and EBV BFLF2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD46497.1"
                     /db_xref="GI:5669990"
                     /translation="MPKSVSSHISLATSTGVRGRSGPRDIRCLSSRLRSYPPGARSASVS
                     SKHRNGLRKFISDKVFFSILSHRIELGVDFLREMETPICTSKTVMLPLDLSTVAPGRC
                     VSLSPFGHSSNMGFQCALCPSTENPTVAQGSRPQTMVGDALKNNELACSVALAFYHHA
                     DKVIQHKTFYLSILSHSMDVVRQSFLQPGLLYANLVLKTFGHDPLPIFTTNNGMLTMC
                     ILFKTRALLHGETALRLLMDNLPNYKISADCCRQSYVVKFVPFTHPDTASIAVQVHTIC
                     EAVAALDCTDEMRDDIQKGTALVNAL"
     polyA_signal   2100..2105
                     /gene="ORF 69"
                     /note="possible signal for ORFs 68 and 69"
     polyA_site   2126
                     /gene="ORF K12"
                     /evidence=experimental
     mRNA   complement(2126..2824)
                     /gene="ORF K12"
                     /product="T0.7"
                     /evidence=experimental
     gene   complement(2126..2824)
                     /gene="ORF K12"
                     /evidence=experimental
     polyA_signal   complement(2139..2144)
                     /gene="ORF K12"

Qy   394 AGGGAGAAAACAGGGACAGCAGCCCATTATCGACACTCTACGGGGACACCCTGGAAGAAACGC   453
             ||:||: |||| | |::|| | ||::|| || || | :| | :| | : ||||||
Db   17861 AGKGWCMYTGSKSKKGKKKIYYASKRGRKKGYTTYWRKMSAVRSRKWWRWAYSMMRCAGMMGAA 17802

Qy   454 AATGCTACCGTGAGTCCCTACAGAACGCCCTTAAACAAGGCAGAGATCTGTGTTCCACC   513
             : :||| :|:|| : || || : || || :|:::: | :|| ::::|:||
Db   17801 CGASRSMKCCWGGYRKSAGSKSSRRGCGYYMKKKGCGGKGSMSSKKWKGSSTSRRGSSAK 17742

Qy   514 CTGAAAAAACAGAGAAGTTCCTGGAGCAGCGGCCAGGATGAGAGACACGGGGAG   573
             : | || |||: | :|| ||: :|| | | || | :||||| | :| |||
Db   17741 SCSYWWGMSMCMCSMMAKSYMCYCYMYRMSSYMSYYKCYSCMGMSSTSYSYSCCWKM 17682

Qy   574 GAGGCCCACCGACTCAAGTGCAAGATGAAAAACCATTGAGACAAATTGAGCTCCTACTCCAG   633
             : || || || :| | | | :| |||:|: | || |::|| ||: |:| |
Db   17741 SSWGSCYKCMKYSGWSSYSTSMGYYSTSCKKYKCSWSSMYKCKTSKSYRRKRSYYWG 17622

Qy   634 AGCCAGCGTTCTGAGGTGGAGGAGACGATTCGAGACACTGGGGACGTCAGCGGTG   693
             : ||| :|| :|||:||||||||| ||| | | || | | || :| |:|||
Db   17621 GGKRAKKKYYCAGRRRMSYWKCCAKWWMSYCCWSYCMTYYYSKSCTYKSSTCYKRGGYY 17562

Qy   694 GAGCAGCTGGCCTGTGTACTGCCGTGTCCCTCAAGAAAGAGTATGAGAATCTGAAGGAAGCT   753
             |:||| |:|| ||| :| | | :|| :||| |: | | | ||:||| |:| | || |:|
Db   17561 WGSKTCYSAGGKSRSMYYCMMRSSKSKSSWSMSMARSSWCMGWGAGYRRSKRSAGWGAGWR 17502

Qy   754 CGGAAGGCCCACAGGGGAACTGGCTGCTGACAGGTTGAAGAAGGATTTGGTGTCCCTAGGAG-   812
             | | | ||| || :||| ||| | :|||| ||| || | | :| :|:| | || |:| ||
Db   17501 SSKGKRSTCMKRACSKKTGSYGSTGRSMKKKKGYSKYSRGMKGRKKTCYCMWKKYKYRKT 17442

Qy   813 CAAGTTGAAGACTCTCAACACTGAGCTGGATCCAGCGCCAAGTTAGAACTGAGGGTCAGCCCA   872
             | ||| |||||| : |:||| ||| ||| | | | |||: :| |||||:| ||| |||||
Db   17441 SMCWWYYMKSWGYKRYRCCMKKKGCTGYRGMSSKKSGYKSMRGMRGSSYSTSCWKSCWG 17382

Qy   873 GAAGGACTTACAAAGTGCTGACCAGGAGATCACGAGCCTAAGAAAGAGAGTCTGATGATCC   932
             | | ||| :| ||||| | | ||||| :|| ||| | |||: || | || |:| |:||
Db   17381 YSWMKCMKYMSYKKKRRSMRGSMSSKGWRGYAGRGCYSSSMWSTRKRRSKCYSYKSYKG 17322

Qy   933 TCCAGG-GAACCTTGAGCCTGCCTCCGCGACCAATGAGACGGTCGAGCGCGCCGGTTTTTG   991
             | | || | |||| | || | |:| :|: | :| | |||: |:||| | ||| |:|:|
Db   17321 RGKMKGWGGMKRGSKYWSSMKKMKRSSNWSKCYSTKSYKSGRRKSKGWGRSTKSRAKSSMR 17262

Qy   992 AGAGCCACGCCCCTGTCGAGATGATGAACCCGAGG--CTTCACCAGCCACCCTTCGGTGA   1049
             | | || | || |:|| | ||:| || ||| || | || ||| || |||| ||:|
Db   17261 MAGSKCTYGSSYWSNNRRNRNMGKTGCNYMYRRSRAMMNGNAAAAGGCTTCCCACANTNGG 17202

Qy   1050 TGAGATTGATCTCAATACCACCCTTTGATCGTAAAATACCCCTCCAACCCAGACCCTCTGGCTC   1109
             | |||| || :| | ::| || :||:| : | ::| |||| ||| |:| ||| |||:
Db   17261 GGAAAAAGGCGGASRASCYKGRMSSKSCYRSGTRRRCMKSKCRRGSRGKSMGMTRGRSG 17142

Qy   1110 CCAGCATTGCCTCCCCAAGAGAAGCTGTGCCGGAGAGGGCACGCTGCTCCCCATGCAGAATGT   1169
             | ||| :|| :|| |||||| |||:|:| :| |||||| | | :|| ||| |:||| ||:
Db   17201 GGAAAAAGGCGGASRASCYKGRMSSKSCYRSGTRRRCMKSKCRRGSRGKSMGMTRGRSG 17142

Qy   1170 CCTCAAGAAGGTGCAACAAAGTCTCCAACCGCGGAGTCCCACGGGTGGCCACGG   1229
             | |||||| :|| | | :|||| | || |:|||| ||| |:| | || |:
Db   17141 GKRTSYSARGSGRGCYCYCWGWGRGRKRCMCSSSRKKMKMYKSYRRKRWMTCMKMCYSMYMAMY 17082

Qy   1230 ATGTGTTAGGAGACTAGAATGAGGCAAACTGGCTGCTGCC   1266
             |:|:||::||||| | ||:||| || ||||||:||| :
Db   17081 CRSMCMCMCMCKSCCGCYSMGMSSYSYSGGKYSWGMKSYWWRSYYSKRSTSKAWRSSKRGMC 17022

Qy   1230 ATGTGTTAGGAGACTAGAATGAGGCAAACTGGCTGCTGCC 1266
             |:|:||::||||| | ||:||| || ||||||:||| :
Db   17021 TGGRYRKGGGRSYGKGGGSYKGGGSWGGKGGGKSTRGSS 16985

```
     CDS             /evidence=experimental
                     complement(2616..2798)
                     /gene="ORF K12"
                     /note="transforming gene"
                     /product="kaposin"
                     /evidence=experimental
                     /codon_start=1
                     /protein_id="AAD46499.1"
                     /db_xref="GI:5669992"
                     /translation="MDRGLTVFVAVHVPDVLLNGWRWRLGAIPPLVCLLAISVVPPSG"

     5'UTR           complement(2799..2824)
                     /gene="ORF K12"
                     /evidence=experimental

     repeat_region   2820..3081
                     /rpt_type=direct
                     /rpt_unit=2820..2842

     repeat_region   3082..3436
                     /rpt_type=direct
                     /rpt_unit=3084..3106

     polyA_signal    complement(6928..6933)
                     /gene="ORF 71"

     CDS             complement(6984..7550)
                     /gene="ORF 71"
                     /note="possible signal for ORF 71"
                     /function="inhibitor of apoptosis"
                     /note="contains two death effector domains"
                     /codon_start=1
                     /evidence=experimental
                     /product="v-FLIP"
                     /protein_id="AAD46498.1"
                     /db_xref="GI:5669891"
                     /translation="MATEVLCEVARKLGTDDREVVLFLINVFIPQPTLAQLIGALRA
                     LKEEGRLTFPLLACLFRAGRRDLLRDLLHLDRFLERHLAGTMSYFSPYQLTVLHVD
                     GELCARDRSLIFLFLSKDTIGSRSTPQTFLHWVCMENLDLLGPTDVDALMSMLRSLSR
                     VDLQRQVQTLMGLHLSGPSHSQHYRHTP"

     polyA_signal    complement(7601..7606)
                     /gene="ORF 72"

     CDS             complement(7601..8406)
                     /gene="ORF 72"
                     /evidence=not_experimental
                     /codon_start=1
                     /note="cyclin D homologue"
                     /product="v-cyclin"
                     /protein_id="AAD46500.1"
                     /translation="MATANNPPSGLLDPTLCEDRIFYNILEIEPRFLTSDSVFGTFQQ
                     SLTSHMRKLLGTWMFSVCQEYNLEPNVVALALNILDRLLIKQVSKEHFQKTGSACLL
                     VASKLRSLTPISTSSLCYAAADSFSRQELIDQEKELLEKLAWRTEAVLATDVTSFILL
                     KLLGGSQHLDFWHHEVNTLITKALVDPKTGSLPASIISAAGCALLVPANVIPQDHSG
                     GVPQLASILGCDVSVLQAAVEQIITSVSDFDLRILDSY"

     gene            complement(7633..8406)
                     /gene="ORF 72"

     gene            complement(7601..12037)
                     /gene="ORF 73"

     CDS             complement(8648..12037)
                     /gene="ORF 73"
                     /note="LNA; component of latency-associated nuclear
                     antigen; LANA"
                     /codon_start=1
                     /evidence=experimental
                     /product="latent nuclear antigen"
                     /protein_id="AAD46501.1"
                     /db_xref="GI:5669894"
                     /translation="MAPPGMRLRSGRSTGAPLTRGSSCRKRNRSPERCDLGDDLHLQPR
                     RKHVADSVDGRECGPHTLPIPGSPPTVFTSGLPAFVSSPTLPVAPITPSPARATPLPPPA
                     LLPPVTISSSPIPPSHPVSPGTTDTHSPSPALPPTQSPESSQRPLSSPTGRPDSSTP
                     MRPPSQQTTPPHSPTTPPEEPPSKSSPDSLAPSTLRSLRKRRLSSPQGPSTLNPICQ"
```

```
     repeat_region   complement(9375..11047)
                     /note="complex array of repeats"

     gene            12624..15262
                     /gene="ORF K14"

     CDS             12624..13670
                     /gene="ORF K14"
                     /note="similar to OX-2 membrane glycoprotein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD46502.1"
                     /db_xref="GI:5669895"
                     /translation="MHITFPGPGRRVVEGGVISSVFLIGAPGRTAIKTEEGVSALQN
                     LPPTVLPVAGTGSVVRRVVCAPPWTTPSASRGSMSSLFISLPWVAFIWLALLGAVGGA
                     RVQGPMRGSAALTCAITPRADIVSTVWQKRQLPGPVNVATYSHSYGVVVQTQYRHKAN
                     ITCPGLMNSTLVIHNLAVDDEGCYLCIFNSPGGRQVSCTACLEVTSPPTGHVQVNSTE
                     DADTVTCLATGRPPPNVTWAAPWNNASSTOEQFTDSDGLTVAMRTVRLPRGDNTTPSE
                     GICLLITWGNESISIPASIQGPLAHDLPAAQGTLAGVAITLVGLFGIFALHHCRRKQGG
                     ASPTSDDMDPLSTQ"

     gene            14111..15139
                     /gene="ORF 74"

     CDS             14111..15139
                     /gene="ORF 74"
                     /note="G protein coupled receptor; similar to IL-8
                     receptor; transforming protein"
                     /codon_start=1
                     /evidence=experimental
                     /product="v-GPCR"
                     /protein_id="AAD46503.1"
                     /db_xref="GI:5669896"
                     /translation="MAAEDFLTIFLDDDESNNETLNMSGYDYSGNFSLEVSVCEMTTV
                     VPYTWNVGILSLFLINVLGNGLVTVIFCKHRSRAGAIDILLLGICLNSLCLSISLLA
                     EVIMLFPNIISTGLCRLEIFFYYLVVLDIFSVVCVSLVRYLVAYSTRSWPKKQSL"
```

```
Query Match          2.5%;  Score 49.6;  DB 14;  Length 28559;
Best Local Similarity  45.6%;  Pred. No. 0.012;
Matches  175;  Conservative  0;  Mismatches  209;  Indels  0;  Gaps  0;

Qy  516  GAAAAAACAGATGAAAGTTCCTGGAGCAGCGGCGATGAGACCAAAACAGCTCGGGAGGA  575
              ||   |            || |  |||||  |||   ||| |  |     ||||| |||
Db 9911  GCAGGAGGAGCAGGAGTTAGAGGAGCCACAGGAGGAGGGNNTLSISOSSQOQOQEPQOQQE 9852

Qy  576  GGCCCACCGACTCAAGTCGCAAGATGAAAACCATGGAGCAAAATTGAGCTCCTACTCCAGAG  635
              || | |  ||||||||  ||||     || ||||  |||| ||  ||| || ||||
Db 9851  GGAGCAGGAGTTAGAGGAGCACAGGAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGA  9792

Qy  636  CCAGCGTTCTGAGGTGGAGCAGATGATTCGAGACATGGGTGTGGGACAGTCACCGGTGGA  695
              || | ||  |||| |||| |||  |||  |   ||||| | |||||  | ||||| ||
Db 9791  GCAGGAGCAGGAGTTAGAGGAGCCACAGGAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGA  9732

Qy  696  GCAGCTGGCTGTGTTACGCGTGTCCCCTCAAGAAAAGACTATATGAAGGAATTCGAAGCTCG  755
              |||| | |||||  || |  ||||| ||   |  ||   |  ||  || || |  |||
Db 9731  GGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGTT  9672

Qy  756  GAAGGCCACACAGGGGAACTGGCTCACAGCTTGAAGAAGGAGATTTGGTTGTCCTCTAGGACCAA  815
              ||| |||   ||  |||  ||  |||  ||| || ||||| |||||| ||  |    ||||
Db 9671  AGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGA  9612
```

```
Qy    816 GTTGAAGACTCTCAAACACTGCAGCTCGGATCAGGCCAAGTTAGACTGAGGTCAGCCCAGA 875
          |||  ||| |||||| |||||||  |||||| ||||||||||||||||||||| |||||||||
Db   9611 GTTAGAGACCAGGAGCAGCGAGCAGTTAGAGGCAGCAGGAGCAGCAGGAGGAGCAGAGGAGGT 9552

Qy    876 GGACTTACAAAGTGCTGACCAGGA 899
          |||||  ||| |||  |||||||||
Db   9551 GGAAGAGCAAGAGCAGGAGGACAGGA 9528
```

```
RESULT 30
LOCUS       AC016837          140207 bp    DNA     linear   HTG 22-DEC-2001
DEFINITION  Homo sapiens clone RP11-5P19, *** SEQUENCING IN PROGRESS ***, 3
            unordered pieces.
ACCESSION   AC016837
VERSION     AC016837.9  GI:17977430
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140207)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome, clone RP11-5P19
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 140207)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL     Direct Submission
            Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 22, 2001 this sequence version replaced gi:16874892.
            All repeats were identified using RepeatMasker;
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L2545
            Center clone name: 5_P_19
            ---------------
            * NOTE:This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *          1      3960: contig of 3960 bp in length
            *       3961      4060: gap of    100 bp
            *       4061      8705: contig of 83645 bp in length
            *       8706     87805: gap of    100 bp
            *      87806    140207: contig of 52402 bp in length
FEATURES             Location/Qualifiers
     source          1..14207
                     /organism="Homo sapiens"
```

```
BASE COUNT    37612 a  29580 c  31494 g  41321 t   200 others
ORIGIN
                     /db_xref="taxon:9606"
                     /clone="RP11-5P19"
                     /clone_lib="RPCI-11 Human Male BAC"
```

```
Query Match           2.5%;  Score 49.2;  DB 2;  Length 140207;
Best Local Similarity 49.7%;  Pred. No. 0.017;
Matches  154;  Conservative    0;  Mismatches 153;  Indels    3;  Gaps    1;

Qy    536 TGGAGCAGCGGCAGGATGAGACCAAAACAAAGCTCGGAGGAGGGCCACCGACTCAAGTGCA 595
          ||||||| |||||||| |||||||||||||||||||||||| |||||| ||||| |||||||
Db  30923 TGGAGCAGCTACAAGATGAGGCGCAAAACACCTGAGGAAGGAGGTGGAGAGTTTGGAGGAA 30982

Qy    596 AGATGAAAAACCATGGAGGACCAAAATTGAGCTCCTACTCCAGAGCCAGCCGTTCTGAGGTGAGG 655
          ||||||| || ||||||||| |||| ||||||||||  |||||||||||||||||||||| |||
Db  30983 AGCTCCAATCCAAGGTGGAAAAACAACCAGGCCGTTGAGTCTCCTGAGCAAGGAACAAAAGT 31042

Qy    656 AGATGAATTCAGACACATGGGTGTGGGACAGTCAGCGTGGGAGCAGTCTGGCTGTGT--ACT 712
          |||||| || ||||||||| ||||||||||||||||||||||||||||||||||  ||  |||
Db  31043 AGAGGCTCCAGAGACCAGGACAGGAGCAGGAGGAGGAGGAGGTGGAGGGTGCGGGAGC 31102

Qy    713 GCGTGTCCCTCAAGAAAAGAGTATGAGAATCTGAAGGAGGCTCGGAAGCGCCACAGGGGAAC 772
          |||||||||  ||||||||||  |||| |||||||||||||||||||| |||||| ||||||
Db  31103 AGGAGAGACTGTGTGACCAAAATGAGAGAGGCTTCGGGAGCAGCAGAAGGAGTTGGGGAGC 31162

Qy    773 TGGCTGACAGTTTGAGAAAGAGATTTGGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACA 832
          |||||||||| |||||||||| |||||||| | |||||||||||||||  |||| |||||||
Db  31163 AGGGTGAGAGGTCGCGAAAGCAAGGAGCAGGAGCGTATCCAAGCAAGGAGGAGGAGGC 31222

Qy    833 CTGAGCTGGA 842
          ||| ||||||
Db  31223 AGGAGGAGGA 31232
```

```
RESULT 31
LOCUS       AC019294          164310 bp    DNA     linear   PRI 02-JAN-2002
DEFINITION  Homo sapiens chromosome , clone RP11-24M17, complete sequence.
ACCESSION   AC019294
VERSION     AC019294.7  GI:17998615
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164310)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome, clone RP11-24M17
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 164310)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Cheepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
JOURNAL     Direct Submission
            Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 164310)
```

```
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
          Brown,A., Camarata,J., Campoppiano,A., Chang,J., Chazaro,B.,
          Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
          Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
          Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
          Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
          Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
          Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (02-JAN-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Dec 29, 2001 this sequence version replaced gi:17977436.
          All repeats were identified using RepeatMasker:
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ----------Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
          ----------Project Information
          Center project name: L4498
          Center clone name: 24_M_17
FEATURES             Location/Qualifiers
     source          1..16431O
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-24M17"
     repeat_region   161..501
                     /rpt_family="MLT1A1"
     repeat_region   936..959
                     /rpt_family="(TTC)n"
     repeat_region   complement(2190..2224)
                     /rpt_family="MER5B"
     repeat_region   complement(2257..2338)
                     /rpt_family="MIR"
     repeat_region   2924..3159
                     /rpt_family="HAL1"
     repeat_region   3464..3762
                     /rpt_family="HAL1"
     repeat_region   3779..3964
                     /rpt_family="AluJb"
     repeat_region   4395..4928
                     /rpt_family="HAL1"
     repeat_region   4985..5237
                     /rpt_family="L1ME3A"
     repeat_region   complement(5439..5653)
                     /rpt_family="L1M4"
     repeat_region   5997..6289
                     /rpt_family="MER20"
     repeat_region   6315..6502
                     /rpt_family="AluSg"
     repeat_region   complement(7097..7401)
                     /rpt_family="AluSg/x"
     repeat_region   7402..7678
                     /rpt_family="AluJb"
     repeat_region   7746..8213
                     /rpt_family="HAL1"

     repeat_region   8249..8726
                     /rpt_family="MLT1H"
     repeat_region   complement(8737..8970)
                     /rpt_family="HAL1"
     repeat_region   8987..9223
                     /rpt_family="MIR"
     repeat_region   9253..9273
                     /rpt_family="HAL1"
     repeat_region   9295..9492
                     /rpt_family="AT_rich"
     repeat_region   10552..10652
                     /rpt_family="HAL1"
     repeat_region   complement(11684..11822)
                     /rpt_family="MIR"
     repeat_region   11891..11966
                     /rpt_family="L2"
     repeat_region   complement(12065..12256)
                     /rpt_family="L2"
     repeat_region   complement(12257..12427)
                     /rpt_family="L1MC4a"
     repeat_region   complement(12841..12957)
                     /rpt_family="L2"
     repeat_region   complement(12981..13274)
                     /rpt_family="FLAM_C"
     repeat_region   13643..13740
                     /rpt_family="AluSx"
     repeat_region   complement(13741..14031)
                     /rpt_family="(CATATA)n"
     repeat_region   complement(14431..14727)
                     /rpt_family="AluJb"
     repeat_region   14752..14972
                     /rpt_family="AluSx"
     repeat_region   15781..16329
                     /rpt_family="L1ME4A"
     repeat_region   16396..16645
                     /rpt_family="MLT1F1"
     repeat_region   complement(16447..16746)
                     /rpt_family="(TA)n"
     repeat_region   17581..17759
                     /rpt_family="AluSg"
     repeat_region   17760..18059
                     /rpt_family="Charlie1"
     repeat_region   18060..18699
                     /rpt_family="AluJb"
     repeat_region   18700..18806
                     /rpt_family="Charlie1"
     repeat_region   18807..19200
                     /rpt_family="FRAM"
     repeat_region   23404..23418
                     /rpt_family="Charlie1"
     repeat_region   23419..23704
                     /rpt_family="L1MC5"
     repeat_region   complement(23705..23304)
                     /rpt_family="AluSx"
     repeat_region   complement(24147..24286)
                     /rpt_family="L1MC5"
     repeat_region   24615..24930
                     /rpt_family="MER20"
     repeat_region   25725..25747
                     /rpt_family="L2"
     repeat_region   complement(26060..26085)
                     /rpt_family="(CCCCG)n"
     repeat_region   26326..26385
                     /rpt_family="GC_rich"
     repeat_region   26639..26660
                     /rpt_family="GC_rich"
     repeat_region   26780..26804
                     /rpt_family="GC_rich"
     repeat_region   26968..27043
                     /rpt_family="GC_rich"
     repeat_region   27469..27543
                     /rpt_family="L2"
     repeat_region   27543..27543
                     /rpt_family="L2"
```

```
repeat_region   27560..27609
                /rpt_family="L2"
repeat_region   27690..27833
                /rpt_family="MER45A"
repeat_region   complement(28662..28762)
                /rpt_family="L2"
repeat_region   complement(28796..29176)
                /rpt_family="MLT1L"
repeat_region   complement(29505..29796)
                /rpt_family="AluSc"
repeat_region   complement(29811..30018)
                /rpt_family="AluY"
repeat_region   complement(30045..30353)
                /rpt_family="AluJb"
repeat_region   30431..30527
                /rpt_family="MIR"
unsure          30539..30559
repeat_region   complement(30969..31248)
                /note="single clone coverage"
```

```
Query Match      2.5%;   Score 49.2;   DB 9;   Length 164310;
Best Local Similarity   49.7%;   Pred. No. 0.017;
Matches 154;   Conservative 0;   Mismatches 153;   Indels 3;   Gaps 1;

Qy    536 TGGAGCAGCGCGGCAGGATGAGACCAAACAAAGCTCGGAGGAGGAGGCCCACCGACTCAAGTGCA 595
          ||||||||| |||||||| |||||||||||| ||| |||||||  |||| ||||||||||||||||
Db  87074 TGGAGCAGCTACAAGATGAGGCCCAAACACCTCAGGAAGGAGGGTGGGAGAGTTTGGAGGGAA 87015

Qy    596 AGATGAAAAACCATGGAGCAAATTGAGCTCCTACTCCAGAGCCAGCCGTTCTGAGGTGGAGG 655
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87014 AGCTCCAATCCAAGGTGGAAAACAACCAGGCCTTGAGTCTCCTGACCAGAGAACAAAAAGT 86955

Qy    656 AGATGATTCGAGACATGGGGTGTGGGACAGTCAGCGGGTGGACGCAGCTGGCTGTTGT---ACT 712
          |||| |||||||  || |  ||||||  ||   ||   ||||  ||||  | ||||| ||| |||
Db  86894 AGGAGAGACTTGTGTGAGCAAAATGACGAGGCTTCGGGAGCAGCACCAGAAAGACGCTACGGGAGC 86835

Qy    713 GCGTGTCCCTCAAGAAAAGAGTATGAGAAATTCTCGGAGGAGGCCACAGGGGGAAC 772
          ||||  | | ||||  |  ||| | || ||||| |||||  || ||||  |||||||
Db  86834 AGGGTGAGAGGCTGCCAAAACGCATGCAACCAGGAGGAGCTGCCAGAGGCTGCGGGAGC 86775

Qy    773 TGGGCTGACAGGTTGAAGAAGGATTTCGTGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACA 832
          |||||| ||| ||||||||||  |||  | ||| | ||| |||| |||||||||| |||||| |
Db  86834 TGGGCTGACAGGTTGAAGAAGGATTTCGTGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACA 832

Qy    833 CTGAGCTGGA 842
          ||| ||||||
Db  86774 AGGAGGAGGA 86765
```

```
RESULT 32
ACCESSION   AC027104/c
LOCUS       AC027104
DEFINITION  Homo sapiens chromosome 15 clone RP11-326L17 map 15, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
                                     180795 bp   DNA   linear   HTG 07-FEB-2002
ACCESSION   AC027104
VERSION     AC027104.5   GI:13490207
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-326L17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
```

```
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,K., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J.T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2001 this sequence version replaced gi:13374714.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------  Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------  Project Information
                Center project name: L8783
                Center clone name: 326_L_17
            -------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1       62900: contig of 62900 bp in length
            *   62901       63000: gap of 100 bp in length
            *   63001      180795: contig of 117795 bp in length
FEATURES             Location/Qualifiers
     source          1..180795
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-326L17"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     55798 a  38910 c  35406 g  50501 t    180 others
ORIGIN
```

```
Query Match      2.5%;   Score 49.2;   DB 2;   Length 180795;
Best Local Similarity   49.7%;   Pred. No. 0.018;
Matches 154;   Conservative 0;   Mismatches 153;   Indels 3;   Gaps 1;

Qy    536 TGGAGCAGCGCGGCAGGATGAGACCAAACAAAGCTCGGAGGAGGAGGCCCACCGACTCAAGTGCA 595
          ||||||||| |||||||| |||||||||||| ||| |||||||  |||| ||||||||||||||||
Db 175486 TGGAGCAGCTACAAGATGAGGCCCAAACACCTCAGGAAGGAGGGTGGGAGAGTTTGGAGGGAA 175427

Qy    596 AGATGAAAAACCATGGAGCAAATTGAGCTCCTACTCCAGAGCCAGCCGTTCTGAGGTGGAGG 655
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175426 AGCTCCAATCCAAGGTGGAAAACAACCAGGCCTTGAGTCTCCTGACCAGAGAACAAAAAGT 175367

Qy    656 AGATGATTCGAGACATGGGGTGTGGGACAGTCAGCGGGTGGACGCAGCTGGCTGTTGT---ACT 712
          |||| |||||||  || |  ||||||  ||   ||   ||||  ||||  | ||||| ||| |||
Db 175366 AGAGGCTCCAGGACCAGGACGACGAGGCTCCGGAGCAGCACCAGAAAGACGCTACGGGAGC 175307

Qy    713 GCGTGTCCCTCAAGAAAAGAGTATGAGAAATTCTCGGAGGAGGCCACAGGGGGAAC 772
          ||||  | | ||||  |  ||| | || ||||| |||||  || ||||  |||||||
Db 175306 AGGAGAGACTGTGTGAGCAAAATGAGAGGCTTCGGGAGCAGCACCAGAGAAGACGCTACGGGAGC 175247
```

```
Qy    773 TGGCTGACAGGTTGAAGAAGGATTTGGTTGGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACA 832
          |||||| |||||| ||| ||||| ||||||| ||||| ||||||| |||| |||| |||| |
Db 175246 AGGGTGACAGGTCGCGAAAGCAGCAGGAGGCTATGCAAGCACAGGAGGAGGCTGCGAA 175187

Qy    833 CTGAGCTGGA 842
          ||| ||| |||
Db 175186 AGGAGGAGGA 175177
```

RESULT 33
AC073564/c

LOCUS   AC073564   227605 bp   DNA   linear   HTG 17-AUG-2001
DEFINITION Mus musculus chromosome 12 clone RP23-32J01 strain C57BL/6J,
            WORKING DRAFT SEQUENCE, 12 unordered pieces.
ACCESSION   AC073564
VERSION   AC073564.3 GI:15208561
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 227605)
  AUTHORS   Smith,D.R.
  TITLE   Genome Therapeutics Corporation Sequencing Center: Mouse Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 227605)
  AUTHORS   Smith,D.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT   On Aug 17, 2001 this sequence version replaced gi:13992669.
            -------------- Genome Center
            Center: Genome Therapeutics Corporation
            Center code: GTC
            Web site: http://www.genomecorp.com/
            Contact: gtc-seqcenter@genomecorp.com
            -------------- Project Information
            Center project name: mg031
            -------------- Summary Statistics
            Sequencing vector: N/A
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 990315
            Consensus quality: 215265 bases at least Q40
            Consensus quality: 216560 bases at least Q30
            Consensus quality: 217418 bases at least Q20
            Insert size: 226604; sum-of-contigs
            Quality coverage: 6.3x in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            ----
            *   1   1184: contig of 1184 bp in length
            * 1185   1284: gap of unknown length
            * 1285   2291: contig of 1007 bp in length
            * 2292   2391: gap of unknown length
            * 2392   3633: contig of 1242 bp in length
            * 3634   3733: gap of unknown length
            * 3734   4891: contig of 1158 bp in length
            * 4892   4991: gap of unknown length
            * 4992   7248: contig of 2257 bp in length
            * 7249   7348: gap of unknown length
            * 7349   10764: contig of 3416 bp in length
            * 10765   10864: gap of unknown length
            * 10865   21712: contig of 10848 bp in length
            * 21713   21812: gap of unknown length
            * 21813   37164: contig of 15352 bp in length
```

FEATURES             Location/Qualifiers
     source          1..227605
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-32J01"
                     /clone_lib="RPCI-23"
     misc_feature    1..1184
                     /note="assembly_name:Contig3"
     misc_feature    1285..2291
                     /note="assembly_name:Contig4"
     misc_feature    2392..3633
                     /note="assembly_name:Contig5"
     misc_feature    3734..4891
                     /note="assembly_name:Contig7"
     misc_feature    4992..7248
                     /note="assembly_name:Contig13"
     misc_feature    7349..10764
                     /note="assembly_name:Contig19"
     misc_feature    10865..21712
                     /note="assembly_name:Contig20"
     misc_feature    21813..37164
                     /note="assembly_name:Contig21"
                     /clone_end:SP6"
     misc_feature    37265..60167
                     /note="assembly_name:Contig22"
     misc_feature    60268..94520
                     /note="assembly_name:Contig23"
     misc_feature    94621..144790
                     /note="assembly_name:Contig24"
     misc_feature    144891..227605
                     /note="assembly_name:Contig25
                     /clone_end:T7"

```
*   37165    37264: gap of unknown length
*   37265    60167: contig of 22903 bp in length
*   60168    60267: gap of unknown length
*   60268    94520: contig of 34253 bp in length
*   94521    94620: gap of unknown length
*   94621   144790: contig of 50170 bp in length
*  144791   144890: gap of unknown length
*  144891   227605: contig of 82715 bp in length.
```

BASE COUNT   69141 a   46827 c   46364 g   63929 t   1344 others
ORIGIN

```
Query Match                2.5%;    Score 49.2;  DB 2;  Length 227605;
Best Local Similarity   44.5%;    Pred. No. 0.018;
Matches  195;  Conservative    0;  Mismatches 243;  Indels    0;  Gaps    0;

Qy    484 TTAAACAAGGCAGAGATGCCTGTGTTCCACCCTGAAAAAACAGATGAAGTTCCTGGAGCAG 543
          |||| ||| |||||||||||| |||| || ||||| | |||| ||||| || |||| |||
Db 185984 TTTAAAAGAGAGTGAAGAAGCTGGACAAGAATGAACAAGAAGAAGAAGAAGAAGAAGAAG 185925

Qy    544 CGGCAGGATGAGACCAAAACAAGCTCGGGAGCAGGAGGCCCACCGACTCAAGTGCAACAA 603
          ||| |||  |||| |||||| ||| | |||||| |||||| || ||||| |||| |||||
Db 185924 AAGAAGAAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 185865

Qy    604 ACCATGGAGCAAATTGAGCTCCTACTCCAGAGCCACCGTTCTGAGGTGGAGGATGGATT 663
          ||| | ||||| |||  ||| |  ||| ||||  ||||| ||||| |||||||| |||||
Db 185864 AAGAGCAGGAAGAAGAAGAAGAAGAAGAAGAGGAAGAGGAAGAAGAAGAAGAAGAAGAAG 185805

Qy    664 CGAGACATGGGTGTGGGACAGTCAGCGGTGGAGCAGCTGGCTGTGTGCTGCTGTCCCTC 723
          | ||||| ||| | ||| |||  ||||  || || | || ||||| || || |
Db 185804 AAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 185745

Qy    724 AAGAAAGAGTATGAGAATCTGAAGGAAAGCTCGGAAGCTCCACAGAGGGGAAACTGGCTGACAGG 783
          || || |||| |  ||| ||||| ||| | | || ||| | ||||| ||||| ||| | ||||
Db 185744 AAGGAGAAGGACAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGGAG 185685

Qy    784 TTGAAGAAGGATTTGGTTGGTGTCCTCTAGGAGCAAGTTGAAGACTCTCAACACTGAGTGGAT 843
          || |  ||| |||||| ||||| ||||| ||| ||| || ||| | ||||| ||| ||| |
Db 185684 AAGGAGAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAG 185625
```

RESULT 34

```
Qy    844 CAGGCCAAGTTAGAACTGAGGTCAGCCCCAGAGAAGGACTTACAAAGTGCTGACCAGGAGATC 903
          |||  |||   |    |||||||||||||||||||||||||||||||| ||  || |||
Db 185624 AAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 185565

Qy    904 ACGAGCCTAAGAAAGAAG 921
          |||  | |  |||||||||
Db 185564 AAGAGAAGAAGAAGAAG 185547
```

LOCUS   AL645746   282611 bp   DNA   linear   HTG 12-JAN-2002
DEFINITION Mus musculus chromosome 13 clone RP23-153B6, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL645746
VERSION   AL645746.2 GI:17148444
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
  AUTHORS   Sims,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Nov 28, 2001 this sequence version replaced gi:17017892.

            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------------- Project Information
            Center project name: bM13B6
            --------------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing program: phrap; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 279809 bases at least Q40
            Consensus quality: 280516 bases at least Q30
            Consensus quality: 281062 bases at least Q20
            Insert size: 281711; sum-of-contigs
            Insert size: 190968; 4.7% error; agarose-fp
            Quality coverage: 12.39x in Q20 bases; sum-of-contigs Quality
            coverage: 18.52x in Q20 bases; agarose-fp
            ----
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES   Location/Qualifiers
     source   1..282611
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-153B6"
                     /clone_lib="RPCI-23"
     misc_feature   1..30613
                     /note="assembly_fragment:06553
     misc_feature   30714..46652
                     /note="assembly_fragment:02685
     misc_feature   46753..51050
                     /note="assembly_fragment:08628
     misc_feature   51151..57360
                     /note="assembly_fragment:06992
     misc_feature   57461..61884
                     /note="assembly_fragment:03821
     misc_feature   61985..71188
                     /note="assembly_fragment:05324
     misc_feature   71289..219059
                     /note="assembly_fragment:05094
     misc_feature   219160..250487
                     /note="assembly_fragment:06258
     misc_feature   250588..257101
                     /note="assembly_fragment:05089"
     misc_feature   257202..282611
                     /note="assembly_fragment:06688"
BASE COUNT   80227 a   60936 c   60357 g   80189 t   902 others
ORIGIN
```

RESULT 35

```
Query Match              2.5%;  Score 49.2;  DB 2;  Length 282611;
Best Local Similarity   45.1%;
Matches  183;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy    516 GAAAAAAACAGATGAAGTTCCTGGACCAGCCGCCGCAGGATGAGACCAAACAAGCTCGGGAGGA 575
          |||||  ||||||||||||||||| |||||||||||||||| |||| |||||||||| |||||
Db 219224 GAAGGAGGAGAAGGAGGAGGAGGAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 219283

Qy    576 GGCCCACGACTCAAGTGCCAAGATGAAAACCATGGAGCAAATTGAGCTCCTACTCCAGAG 635
          |||| || | | |||   |||| || |||   | ||| |||  ||||||| |||||||||
Db 219284 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGA 219343

Qy    636 CCAGCGTTCTCGAGGTGCGAGGAGATGATTCGAGACATGGGTGTGTGGGACAGTGCAGGTGGA 695
          ||||    |||  | | |||||||| | || |||||  |  |||| |  | |||  |   |||
Db 219344 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 219403

Qy    696 GCAGCTCGGCTGTGTACTGCGGTGTCCCTCAGAAAGGTATGAGAATCTGAAGGAAGCTCG 755
          ||||  | | |  | ||   | |  |   |||  | ||| |||| || |||||||||||| 
Db 219404 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 219463

Qy    756 GAAGGCCACAGGGGAACTGGCTGACCAGGTTGAAGAAGGAGATTTGGTGTTCCCTCTAGGAGCAA 815
          ||||  | |||   ||   || || ||||||||||  |||| || ||| || |||   |||| 
Db 219464 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 219523

Qy    816 GTTGAAGACTCTCAACACTGAGCTGATCAGGCCAAGTTACAAGTTAGAACTGAGGTCAGCCCCAGAA 875
          ||||| |   | |   || |||  | |||||| |||||||||||| |||||||||| |   || 
Db 219524 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 219583

Qy    876 GGACTTACAAAGTGCTGACCAGGAGATCACGAGCCTAAGAAAGAAG 921
          ||  |||||||| | | ||| ||||  ||||| || |||||||||
Db 219584 GAAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 219629
```

LOCUS   AC084389   206497 bp   DNA   linear   ROD 26-SEP-2001
DEFINITION Mus musculus clone RP23-247A1, complete sequence.
ACCESSION   AC084389
VERSION   AC084389.1 GI:15778822
KEYWORDS   HTG.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 206497)
  AUTHORS   McPherson,J.D.
  TITLE   The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 206497)
  AUTHORS   McPherson,J.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, Missouri 63108, USA
COMMENT   Center project name: M_BA0247A01.
FEATURES   Location/Qualifiers
```

```
RESULT 36

LOCUS       AY033497               2160 bp    mRNA    linear   ROD 01-JAN-2002
DEFINITION  Mus musculus keratin protein K6irs mRNA, complete cds.
ACCESSION   AY033497
VERSION     AY033497.1  GI:18031723
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2160)
  AUTHORS   Porter,R.M. and McLean,W.
  TITLE     Keratin K6irs is specific to the inner root sheath of hair
            follicles in mouse and humans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2160)
  AUTHORS   McLean,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Human Genetics Unit, University of Dundee,
            Ninewells Hospital and Medical School, Dundee DD1 9SY, UK
FEATURES             Location/Qualifiers
     source          1..2160
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
     5'UTR           1..51
                     /note="MK6irs.p1"
     primer_bind     1..20
     CDS             52..1626
                     /note="type II intermediate filament protein; expressed in
                     the inner root sheath of hair follicles"
                     /codon_start=1
                     /product="keratin protein K6irs"
                     /protein_id="AAK55510.1"
                     /db_xref="GI:18031724"
                     /translation="MSRQFTCKSGASNRGFSGCSAVLSGGSSSSSYRAGGKGLSGGFGS
                     RSLYSLGGGRSITLLNMASGSGKNGFGFGRNRASGFAGSIFGSVALGPVCPAVCPGG
                     IHQVTVNESLLAPLNVELDPEIQKVRAQERQIKALLNNKEFASFIDKVRFLEQQNQVLQ
                     TKWELLQQLDLNNCKNNLEPLLEGHISNMRQLETLSGDRVRLDSELRNVRDVVEDYK
                     KKYEEEINRRTAAAENEFVLLKKDVDAAYAAKVELQAKVDVDMQDIKFFCLFEAEMAQ
                     IQSHISDMSVILSMDNNRNLDLDSIIDEVRAQYEIIALKSKAAEEALYQTKFQEQLA
                     AGRHGDDLKNTKNEITELTRFIQRLRSEIENAKKQASNLEFIAIADAEQRGDSALRDAR
                     AKLDELEGALHQAKEELARMLREYQELMSLKLALDMEIATYRKLLESEECRMSGEYSS
                     PVSISTISSTSSGSGYGFPRSTVSGGIVANSTSCISGVCSVRGGENRSRGSASDYKDT
                     LTKGSSLSTPSKKGGR"
     3'UTR           1627..2160
                     /note="MK6irs.p2"
     primer_bind     complement(2141..2160)
     polyA_signal    2156..2160
ORIGIN

BASE COUNT    503 a    603 c    609 g    445 t
```

```
Query Match    2.5%;  Score 48.8;  DB 10;  Length 2160;
Best Local Similarity  47.2%;  Pred. No. 0.018;  Matches 167;  Conservative  0;  Mismatches 187;  Indels  0;  Gaps  0;

Qy  527 TGAAGTTCCTGAGCACGCGGCAGGATGAGACCAAAACAAAGCTCGGGAGGAGGGCCCACGAC 586
        ||||||||||  ||||| || |||| | |||  | || |  | ||| | || || ||||||
Db  506 TGGACAGCAGAACCAGGTGCTGCAGACCAAGTGCGCCAGCCTGCTCGCGCCAGCTGGACCTGA 565

Qy  587 TCAAGTGCCAAGATGAAAAACCATGGAGCAAAATGAGCTCCTACTCCAGAGTTCTG 646
        ||||||||||| ||| |||   | |  | | |   | || || || |||| |||||
Db  566 ACAACTGCCAAGAAACAACCTGAACCATCCTTGAGGGCCACCATCACCAACATGCGGAAGC 625

Qy  647 AGGTGGAGGAGATGATTCGAGACATGGGTGTGTGGGACAGCACGGTGGAGCAGCTGGCTG 706
        ||  ||  |  | || ||| ||  | |||  |   || || || |||| || || || ||
Db  626 AGCTGGAGACGCTGTCTGGGGACAGAGTGAGGCTGGACTCGGAGCTGAGGAATGTGCGTG 685

Qy  707 TGTACTCCGTGTCCCTCAAGAAAGAAGTATGAAGATCTGAAGAGAGGCTTC 766
        |||||||   ||| | ||| |||| || |  | | ||  | |||| | |||
Db  686 ACGTGGTGGAGGACTACAAGAAGAAGTATGAGGAGGAGATCAACCGCCGGACAGCTGCAG 745

Qy  767 GGGAACTGGCTGACAGGTTGAAGAAGGATTTGGGTGTCCTCTAGGAGCCAAGTTGAAGACTC 826
        | | ||||| || | |||    ||||||| || |  |   |||||| | ||||| | ||
Db  746 AGAATGAGTTGTTGCTCGTCCTGAAGAAGGACGTGGATGCGGCTTATGCCAACAAGGTGGAAC 805

Qy  827 TCAACACTGAGCTGGA 842
        | |  ||| || ||||
Db  806 TGCAGGCCAAGGTGGA 821
```

```
RESULT 37

LOCUS       AB033744               2193 bp    mRNA    linear   ROD 22-OCT-1999
DEFINITION  Mus musculus mRNA for type II cytokeratin, complete cds.
ACCESSION   AB033744
VERSION     AB033744.1  GI:6092074
KEYWORDS    type II cytokeratin.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Aoki,N. and Tsujimoto,T.
  TITLE     Isolation and characterization of a novel mouse type II cytokeratin
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 2193)
  AUTHORS   Aoki,N. and Tsujimoto,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1999) Noriaki Aoki, Niigata University School of
            Medicine, Department of Dermatology; 1 Asahimachi-dori, Niigata
            951-8510, Japan (E-mail:rayaoki@med.niigata-u.ac.jp.
```

```
source          1..206497
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="m"
                /clone="RP23-247A1"

BASE COUNT    52679 a    51024 c    48836 g    53958 t
ORIGIN

Query Match    2.5%;  Score 49;  DB 10;  Length 206497;
Best Local Similarity  45.6%;  Pred. No. 0.02;  Matches 172;  Conservative  0;  Mismatches 205;  Indels  0;  Gaps  0;

Qy  524 AGATGAAGTTCCTGAGCACGCGGCAGGATGAGACCAAACAAAGCTCGGGAGGAGGGCCCACC 583
        |||||||||||||||| ||||| |||| || | || |  |||| ||||||||||||||||||
Db 76994 AGGAGCAGGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGA 76935

Qy  584 GACTCAAGTGCATAGATGAAAAACCATGGAGCAAAATTGAGCTCCTACTCCAGAGGAGCGTT 643
        ||| |  |||   || |||| |  ||  |||| | |  ||| || || | |  |||| ||
Db 76934 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGA 76875

Qy  644 CTGAGGTGGAGGAGATGATTCGAGACATGGGTGTGTGGGACAGCAGCTACGGTGTGGAGCTGG 703
        |||||||||| || ||| | || || | |||| |   || ||| ||  | |||| || || |
Db 76874 ATGGAGGAGGAGCAGGAGGAGGAGGAGGACAGGAGGAGCAGGAGGAGGAGGAGGAG 76815

Qy  704 CTGTGTACTCCGTGTCCCTCAAGAAAGAGTATGAGAATCTGAAGAAGCTCGGGAGGAGGCCA 763
        ||  |||||   ||| | ||| ||||| | |  ||| ||  |||| ||| || || || |||
Db 76814 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGAGCAGGAGGAGGAGCAGG 76755

Qy  764 CAGGGGAACTGGCTGACAGGTTGAAGAAGGATTTGGTGTCCTCTAGGAGCAGAGTTGAAGA 823
        | ||  ||||| || | ||    |||||||| | || | |   |||||| || |||| | |
Db 76754 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGG 76695

Qy  824 CTCTCAAACACTGAGCTGGATCAGGCCAAGTTAGAACTGAGCTGGAGGAGGACTTAC 883
        | | |  || ||| || ||||  ||| ||  |   || || || || |  ||||
Db 76694 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 76635

Qy  884 AAAGTGCTGACCAGGAG 900
        ||| | || ||||||||
Db 76634 AGGAGGAGGAGCAGGAG 76618
```

FEATURES   Tel:81-25-227-2282, Fax:81-25-227-0783)
    source   Location/Qualifiers
                  1..2193
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
    CDS   52..1626
                  /codon_start=1
                  /product="type II cytokeratin"
                  /protein_id="BAA85657.1"
                  /db_xref="GI:6092075"
                  /translation="MSRQFTCKSGASNRGFSGCSAVLSGGSSSSYRAGGKGLSGGFGS
                  RSLYSLGGGRSITLNMASGSGKNGEGFGRNRASGFAGSIFGSVALGPVCPAVCPGG
                  IHQVTVNESLLAPLNVELDPEIQKVRAQEREQIKALNNKFASFIDKVRFLEQQNQVLQ
                  TKWELLQQLDLNNCKNNLEPTLEGHISNMRKQLETLSGDRVRLDSELRNVRDVVEDYK
                  KKYEEEINRTAAENEFVLLKKDVDAAYANKVELQAKVDTMDQDIKFFKCLFEAEMAQ
                  IQSHISDMSVILSMDNNRNLDLSIIDEVRAQYEEIALKSKAEAEALYQTKFQELQLA
                  AGRHGDDLKNTKNEITELTRFIQRLRSEIENAKKQASNLETAIADAEQRDGSALKDAR
                  AKLDELEGALHQAREELAKMLREYQEELMLKLALDMEIATYRKLLESEECRMSGEYSS
                  PVSIISSTSSGGYGFRPSTVSSGGYVANSTCISGVCSVRGGENRSGASDYKDT
                  LTKGSSLSTPSKKGGR"
    polya_signal 2156..2161
BASE COUNT   528 a   607 c   611 g   447 t
ORIGIN

Query Match   2.5%;   Score 48.8;   DB 10;   Length 2193;
Best Local Similarity   47.2%;   Pred. No. 0.018;
Matches 149;   Conservative 0;   Mismatches 167;   Indels 0;   Gaps 0;

```
Qy  527 TGAAGTTCCTCGGAGCAGCGCGGCAGGATGGAGAGACCAAACAAGCTCGGGAGGAGGAGCGCCACCGAC 586
        || ||  |||  ||||||||
Db  506 TGGAGCAGCAGAACCAGGTGCTCGCACACAAGCTGGGAGCTCCAGCAGCTGGACCTGA 565

Qy  587 TCAAGTGCAAGATGAAAACCATGGACAAATTGAGCTCCTACTCAGGAGCCAGCGGTTCTG 646
        ||| |||||| ||||||  |  |||
Db  566 ACAACTGCAAGAACAACCTGGAACCCATCCTTGAGGGCCACATCAGCAACATGCGGAAGC 625

Qy  647 AGGTGGAGGAGATGATTCGACAACATGGTGTGGGAACGTCAGCGGCAGCTGGCCTG 706
        ||||||  ||||  ||||  ||  |  |||
Db  626 AGCTGGAGACCCTGTCTGGGGACAGAGTGAGGCTCGACTCTGAGCTGAGGAATGTCCGTG 685

Qy  707 TGTACTGCGGTGTCCCTCAAGAAAAGGAGTATGAGAATCTGAAGGAAGGCCCACACG 766
        ||||  |  ||||  ||  ||||
Db  686 ACGTGGTGGAGGACTACAAGAAGAAGTATGAGGAGGAGATTAATCGGACAGCTGCAG 745

Qy  767 GGGAACTGGCTGGACAGGTTGAAGAAGGAGATTTGGTGTCCTCTAGGAGACAAGTTGAAGACTC 826
        || |||||||  |||||||  ||||
Db  746 AGAATGAGTTTGTGCTGCTGAAGAAGGACGTGGATGCGGCTTATGCCAACAAGGTGGAAC 805

Qy  827 TCAACAACTGACCTGGA 842
        ||| ||
Db  806 TGCAGGCCAAGGTGGA 821
```

RESULT 38

ACMHC
LOCUS   ACMHC   5894 bp   DNA   linear   INV 10-FEB-1999
DEFINITION   Acanthamoeba castellanii nonmuscle myosin heavy chain gene.
ACCESSION   Y00624 M12702 M12703 M19549
VERSION   Y00624.1 GI:5585
KEYWORDS   myosin heavy chain.
SOURCE   Acanthamoeba castellanii
  ORGANISM   Acanthamoeba castellanii
               Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE   1 (bases 1 to 5894)
  AUTHORS   Hammer III,J.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-OCT-1987) Hammer III,J.A., Lab. of Cell Biology, BLD6
               3, RM BI-22, National Institutes of Health, Bethesda, MD 20892
REFERENCE   2 (bases 1 to 5894)
  AUTHORS   Hammer,J.A. III, Bowers,B., Paterson,B.M. and Korn,E.D.
  TITLE   Complete nucleotide sequence and deduced polypeptide sequence of a
               nonmuscle myosin heavy chain gene from Acanthamoeba; evidence of a

  JOURNAL   hinge in the rodlike tail
  MEDLINE   J. Cell Biol. 105 (2), 913-925 (1987)
               87308395

FEATURES   Location/Qualifiers
    source   1..5894
                  /organism="Acanthamoeba castellanii"
                  /db_xref="taxon:5755"
                  /clone_lib="lambda 2001"
    precursor_RNA 517..>517
                  /note="minor primary transcript"
    precursor_RNA 560..>560
                  /note="major primary transcript"
    mRNA   join(560..895,1310,1381..1565,1743..>1743)
    CDS   join(689..895,1161,1310,1381..1565,1743..>5730)
                  /codon_start=1
                  /product="myosin heavy chain"
                  /protein_id="CAA68663.1"
                  /db_xref="GI:5586"
                  /db_xref="SWISS-PROT:P05659"
                  /translation="MAAQRRRGGEVESDYIKYLKYKNTGFQVSASDKTLAWWPIKDA
                  DRAFCHVEVIKDDGKNFTVRLENGEEKSOPKNEKNFLGVNPPKFDGVEDMGELGYLNE
                  PAVLHNLKRYDADLPFHTYSGLFLVVNPAYRPLPVYTPEIIDIYRQRQRDKVAPHIFA
                  ISDAAYRAMLNTRQNQSMLITGESGAGKTENTKKVTCIQYLTAIAGRAEGGLLEQQILFA
                  NPILEAFGNAKTTKNNNSSRFGKFIELQFNANGGQITGANIFIYLLEKSRVTAQGAGER
                  NFHIFYQILSKAMPELKQKIKLTKPEDYFLNQNACYTVDDMDDAKEFDHLKRAFDI
                  LNINEERLALFQTISAILHLGNLPFIDVNSETAGLKDEVELNIAAELLGVSAAGLKA
                  GLLSPRIKAGNEWVTRALNKPKAMASRDALCKALFGRLFLMIVQKINRILSHKDTAL
                  WIGVLDISGFEIFQHNSFEQLCINTFNELQQFNHMFLEQQEYEREEKIIDWTEYDV
                  GMDSQPCIDLIEKKPMGILPLLDEOYVPPADDTSPTKLKLFQTHNHRNRRPPRPDAN
                  NFKIVHYAGEVEYQISAWLEKNRDPLEDDLSNLCKKSSVRFVTGLFPEDLMPSFRAAP
                  AEEEKAAAGGSRNRSTGRGRKGGAOFITVAFQYKEQLALHLMSLSSTAPHFIRCIIPNL
                  GKPGVVSDQLVLDQLKCNGVLEGIRIARGWPNRLKYDEFLKRYRFLLKPGATFPSPS
                  TKDAVKDLIEHLIAKEPTKVNKDEVRFGVKIFFRSGQLAAIEELREQAISKMVVSIQ
                  AGARAFLARRMYDKMREQTVSAKILQRNIRAWLELKNWAWTQLYVRARPLISQRFEQK
                  EIDDLKKQVKDLEKELAALKDANAKLDKEKQLAEEDADKLKDLAALKILDLEGEK
                  ADLEEDNALLQOKKVAGLEELLQEETSASNDILEORKLEABKGELKASLEEEDRNRKA
                  LQEAKTKVEESERNELODKYEDEAAAAHDSLKKKEEDLSRELRETKDALADAENISETLR
                  SKLKNTERGADDVRNELDDVTATKLOLEKTKKSLEEEETQRAQLEEKSGKEAASSK
                  AKQLQOLEDARSEVDSLKSKLSAAEKSLKTKADQNRDDELQEDERTVRANDVKQKK
                  ALEAKLTELEDQVTALDGQKNAAAAQAKTLKTQVDETKRRLEEAANAAAARNELEENA
                  LDEVAQLTADLDAERDSGAQQRRKLNTRISELQSELENAPKTGGASSEEVKRLEGELE
                  RLEEELLTAQEARAAAEKNLDKANIELEELRQEADAAAORDKLVKDNRLKLKADLDEA
                  RIQLEEEQDAKSHADSSSRRILAEIELEKRKVVAKETSDQKOAAEDQORKNSEL
                  DRDSIERRNRDAERQVRDLRAQLDDALSRLDSEKRAKEKSVEANRELKKVVLDREKQS
                  LESLSKFNSALESDKQILEDEIGDLHEKNLQAKIAQLQDEIDGTPSSRGGSTRGAS
                  ARGASVRAGSARAEE"
    intron   896..1160
                  /note="intron I"
    intron   1311..1380
                  /note="intron II"
    intron   1566..1742
                  /note="intron III"
    misc_feature 5799..5804
                  /note="pot. polyA signal"
BASE COUNT   1311 a   1980 c   1615 g   988 t
ORIGIN

Query Match   2.5%;   Score 48.8;   DB 3;   Length 5894;
Best Local Similarity   44.2%;   Pred. No. 0.019;
Matches 246;   Conservative 0;   Mismatches 307;   Indels 3;   Gaps 1;

```
Qy  396 GGAGAAACGGGACAGCCAGGCCCATTATCGCACTCTACGGGACACCTGGAAGAACGCCAA 455
        |||||| ||  |||||||  ||   ||  ||||||  |||||
Db 4041 GCAGAACGGCAAAGCTCGAGGCCGAGAAGGGCCAGCTCAAGGCCTCCGTCGAGGAGGA 4100

Qy  456 TGCTACCGTGGAGTCCCTACAGAACCGCCTTAAACAAGGCAGAGATGCGTGTTCACCCT 515
        || ||||| ||||||||  |||  ||||   ||  ||||||||
Db 4101 GCGCAACCGCAAGGCCCTCCAGGAGAGCCAAGACCAAGGTCGAGTCCGACGAGTT 4160

Qy  516 GAAAAAACAGAGATGAAGTTCCTGGACAAGACCAGACCAAACAAGCTCGGGAGGA 575
        |  |  ||||| |||||  ||  ||||   ||   |||||||||||||  |||||
Db 4161 GCAGGACAAGTACGAGGACGCCGCGACAAGACCAAGCTCGGCAGACAGAAGAGGAGGA 4220
```

```
RESULT 39
LOCUS       AL589661
DEFINITION  Mouse DNA sequence from clone RP23-58B7 on chromosome 15.
ACCESSION   AL589661
VERSION     AL589661.21  GI:18476659
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 241432)
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18151481.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations,
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-58B7 is
            from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-58B7.
FEATURES             Location/Qualifiers
     source          1..241432
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP23-58B7"
                     /clone_lib="RPCI-23"
     misc_feature    147938..148190
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    64293 a  50995 c  55888 g  70256 t
ORIGIN
```

```
Qy    576  ---GGCCCACCGACTCAAGTGCAAGATGAAAACCATGGAGCAAAATTGACTCCTACTCCA  632
           |  ||| ||| ||  |   || ||||||  | ||||||| |||  |||||||||||||||
Db   4221  CCTGTCGCCGCGAGCTTCCGCGAGACCAAGGACGCCCCTGGCCGCCGAGAACACTCTCGA  4280

Qy    633  GAGCCAGGCGTTCTGAGGTGGGAGGAGATGAATTCGAGACATGGGTGTGGGACAGTCGGT  692
           ||||  ||| || ||   || |||||| |  ||||  ||| || || |||| || |||||
Db   4281  GACCCTGGCTCCTCAAGCTCAAGAACACCACGGACGCGGCGCGCGACGACGTCGGCAGCT  4340

Qy    693  GGAGCAGCTGGCTGTGTGGTACTGCGGTGTCCCTCAAGAAAAGAGTATGAGAATCTGAAG  752
           ||| || || ||  |  ||   |||  |   |||| ||||||| |||| ||| |  ||| |
Db   4341  CGACGACGTCACCGCCCCACCAAGCTCCACGTTGAGAAGAGACCAAGAAGAGCCTCGAGGAGGA  4400

Qy    753  TCGGAAGGCCACAGGGGAACTGGCTGACACGTTGAAGAAGGATTTGGGTCGTCCTCTAGGAG  812
           ||||| ||||   ||||||  ||||| |||||||||||||| |||||| || |||||||||
Db   4401  GCTCGCCCCAGACCCCGCGCCCAGCTCGAGGAGGAGGAGCGCCAAGGAGGCGCCCCAGCTC  4460

Qy    813  CAAGTTGAAGACTCTCAACACTGACTGGATCAGGCCAAGTTAGAACTGAGGTCAGCCCA   872
           || |||||||| || ||   |||  |||| ||| || ||||  ||||| || |||||||
Db   4461  CAAGCCAAGCAGTCTGGCGCCGGCACCGCCTCCAGGAGGAGGAGGAGTCGCACTACACTCTAC  4520

Qy    873  GAAGGACTTACAAAGTGCTGACCAGGAGAATCACCAGGACGGTAAGAAAGATCTGATGATCC  932
           |||||||      |||  |||||| |  ||      | ||  ||  | ||  |||  ||||
Db   4521  GAGCAAGCTCTCGGCGCCGGCCGAAGTCGCCAAGGACCAGAACCGGGACCAGGAGGGGACCT  4580

Qy    933  TCCAGGGAACCTTGAG  948
           | |||||| |||||||
Db   4581  CGACGCAGCAGCTTGAG  4596
```

```
RESULT 40
LOCUS       AC016961
DEFINITION  Homo sapiens chromosome 3 clone RP11-394J21, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC016961
VERSION     AC016961.16  GI:17530684
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202010)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
```

```
LOCUS       AL589661            241432 bp    DNA     linear   ROD 30-JAN-2002

Query Match           2.4%;  Score 48.2;  DB 10;  Length 244132;
Best Local Similarity  42.8%;  Pred. No. 0.034;
Matches  242;  Conservative   0;  Mismatches 323;  Indels   0;  Gaps   0;

Qy    314  AGGAGGAGGAGAATGCTGTTGGATGCAGAATTCTTAAAGAATGAACTGGACAGCGTCAAAG  373
           ||||||||| || ||||| ||||| ||  ||||| ||||| |||| ||||||| |||||||
Db  147794  AGGAGGAACAGGAGGAGGAACAGGAGGAACAGGAGGAGGAGGAACCATTATCGACACTCTAC  147853

Qy    374  CTCAGCTTTCCCAGAAAGACAGGAGGAGAAACCGGACAGCCAGTTATCGACACTCTAC  433
           || || ||  |  ||||| ||||||||    || ||||| ||| || |  ||| |||
Db  147854  AACAGGAGGAGGAGGAACAGGAGGAGGAGCAGGAGCAGCCAGGAGGAGGAGGAACAGG  147913

Qy    434  GGGACACCCTGGAAGAACGCAATGCTTACCGTGTGGAGTCCCTACAGAACGCCTTTAAACAAG  493
           ||||  | ||||||  || | |  | |||  || ||||  || ||||  |||| | |||||
Db  147914  AGGAGGAACAGGAGGAACAGGAGGAGGAACAGGAGCAGCCAGGAGGAGGAACAGG  147973

Qy    494  CAGAGATGCTCGTGTTCCACCCTGAAAAAAACAGATGAAGTTCCTGGACGAGGCAGGATG  553
           || ||||||| || ||  || |||| |||||||||||||  ||||||| |||| | | ||
Db  147974  AGGAGGAGGAGCAGGAGGAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGCAGGAGG  148033

Qy    554  AGACCAAAACAAGCTCGGGAGGAGGAGGCCCACCGGACTCAAGGTGCAAGATGAAAACCATGGAGC  613
           |||| ||||| ||||  ||||||||| ||    | | |||||||||||||| ||||| |  ||||
Db  148034  AAGAGGAGGAGGAGGACAGGAGGAGGAACAGGAGGAGGAACAGGAGGAGGAGGAGGAGGAACAGG  148093

Qy    614  AAAATTGAGCTCTACTCCACGAGCCACGTTCTGAGGTGGGAGGAGATGAATTCGAGACATGG  673
           ||||||||  | ||  || ||||| |  || || ||| ||||||| |  ||||  ||| ||
Db  148094  AGGAGGAACAGGAGGAGGACAGGAGGAGGAGGAACAGGAGGAGGAGGAACAGGAGGAGG  148153

Qy    674  GTGTGGGACAGTCAGGCGGTCGAGCAGCTGGGTGTGTGTTGTACTGCGGTGTCCCTCAAGAAAAGAGT  733
           | |||||| || |||| ||  ||||| |  ||   |  ||  |||  || |   ||||  || |  ||
Db  148154  GGCAGGAGGAGCAGGAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAACAGGAGGAACAGG  148213

Qy    734  ATGAGAANTCTGAAGGAAGCTCGGAGCGCCACAGGGGAACTGGCTGACAGGTTGAAGAAGG  793
           |||||| ||||||| ||||  |||||  |  ||||||||  ||||| |||||||||||||
Db  148214  AAGAGGAACAGGAGGAACAGGAGGAGGAACAGGAGGAGGAGGAGCAGG  148273

Qy    794  ATTTTGTGTCTCTCTAGGAGCAAGTTGAAGACTTCTCAACACTGACTGAGCGTGGATCAGGCCAAGT  853
           |  | ||  | ||  || ||| |||| | | || || ||| |||| ||||||  |||| |||| ||
Db  148274  AGGAGGAGCAGGAGGAGGCGCAGGAGGAGGAGGAACAGGAGGAGGAGGAGGAACAGG  148333

Qy    854  TAGAACTGAGGTCAGCCCAGGAAGGA  878
           || |||||||| |||| |||||| ||
Db  148334  AGGAGGAGCAGGAGGAGGGGGAGGA  148358
```

```
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
             Burch,P., Burkett,K.L., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
             Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
             Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
             Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
             Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
             Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
             Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
             Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
             Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
             Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newtson,N.,
             Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
             Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
             Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
             Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
             Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
             Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
             Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
             Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
             Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
             Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
             Watlington,S., Williams,G., Williamson,A., Wieczyk,R., Wooden,S.,
             Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2 (bases 1 to 202010)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
             On Dec 12, 2001 this sequence version replaced gi:17432834.
COMMENT      -------------- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             -------------- Project Information
             Center project name: HMVB
             Center clone name: RP11-394J21
             -------------- Summary Statistics
             Sequencing vector: M13; L08821
             Chemistry: Dye-primer Bodipy: 2% of reads
             Chemistry: Dye-terminator Big Dye: 98% of reads
             Assembly program: Phrap; version 0.990329
             Consensus quality: 204811 bases at least Q40
             Consensus quality: 206552 bases at least Q30
             Consensus quality: 207618 bases at least Q20
             Estimated insert size: 203982: sum-of-contigs estimation
             Quality coverage: 8.1x in Q20 bases: sum-of-contigs estimation
             --------------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             *   consists of 6 contigs. The true order of the pieces
             *   is not known and their order in this sequence record is
             *   arbitrary. Gaps between the contigs are represented as
             *   runs of N, but the exact sizes of the gaps are unknown.
             *   This record will be updated with the finished sequence
             *   as soon as it is available and the accession number will
             *   be preserved.
             * be preserved.
             *                     1 143460: contig of 143460 bp in length
             *                143461 143560: gap of unknown length
             *                143561 164598: contig of 21038 bp in length
             *                164599 164698: gap of unknown length
             *                164699 178504: contig of 13806 bp in length
             *                178505 178604: gap of unknown length
             *                178605 188644: contig of 10040 bp in length
             *                188645 188744: gap of unknown length
             *                188745 198961: contig of 10217 bp in length
             *                198962 199061: gap of unknown length
             *                199062 202010: contig of 2949 bp in length.
FEATURES             Location/Qualifiers
     source          1..202010
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-394J21"
BASE COUNT     59923 a   43696 c   42411 g   55465 t    515 others
ORIGIN
```

```
Query Match         2.4%;    Score 47.4;   DB 2;   Length 202010;
Best Local Similarity 45.7%;   Pred. No. 0.058;
Matches  165;  Conservative   0;  Mismatches 196;  Indels   0;  Gaps   0;

Qy  520 AAACAGATGAAGTTCCTGGAGCAGACGCGGCAGGATGAGACCAAACAAGCTCGGGAGGAGGCC 579
            |||||||||||||  ||||||| ||  ||||||||||  |||| ||||| |||||||||||||
Db  201627 AAAAAAAAAAAAAAAAAAGAGGAGGAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 201686

Qy  580 CACCGGACTCAAGTGCAAGAGATGAAAAACCATGGAGCAAAATTGAGCTCCTTACTCCAGACCCAG 639
            |||  |||  ||||| ||||||  |||||||||||||||  ||||||       |||  ||
Db  201687 GAGGAGGAGGAGGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAG 201746

Qy  640 CGTTCTGAGGTGGAGGAGGATGATTCGAGACATGGGTTGTGGGACAGTGACGCGGTGGGAGCAG 699
            ||||  |||||||||||||||| | |||   ||  ||   ||  || |||    || |||
Db  201747 AAGGAGGAGGAGGAGGAGGAGGAGGACGAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 201806

Qy  700 CTGGCTGTGTACTGCGTGTCCCTCAAGAAAGAGTATGAGAATCTGAAGGAAGCTCGGAAG 759
            |||| ||  ||  ||||||| | ||||||   ||  |||||  |||||||| |||||||||
Db  201807 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 201866

Qy  760 GCCACAGGGGAACTGGCTGACAGGTTGAAGAAGGAGATTTGGTTGTCCTCTAGGAGCAAGTTG 819
            ||| | |||||  |||| | |||| ||||| |  || ||  || |||| |||||||||||||
Db  201867 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAGAGAAG 201926

Qy  820 AAGACTCTCAACACTCGAGCTGGATCAGGCCAAGTTAGAACTGAGGTCAGCGCCAGGAAGGAC 879
            ||||         | |||| |||| |||||||||| |||   || |||      |||
Db  201927 AAGAGAGGAGAGAGGAGGAGGAGGAGGAGGAAGGAGGAGGAGGAGTTGAGACCGCCAGGAGGAG 201986

Qy  880 T 880
            |
Db  201987 T 201987

Search completed: September 4, 2002, 15:18:52
Job time: 18391 sec
```

THIS PAGE BLANK (USPTO)